

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic ~ nucleic search, using sw model

Run on: April 3, 2004, 05:12:10 ; Search time 5955 Seconds
(without alignments)
11332.511 Million cell updates/sec

Title: US-10-029-180-7
Perfect score: 1557
Sequence: 1 atggatcctagaacacatcc.....cogttcagccgccgagacaa 1557

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*

- 1: gb_ba.*
- 2: gb_htg.*
- 3: gb_in.*
- 4: gb_om.*
- 5: gb_ov.*
- 6: gb_pat.*
- 7: gb_ph.*
- 8: gb_pl.*
- 9: gb_pr.*
- 10: gb_ro.*
- 11: gb_sts.*
- 12: gb_sy.*
- 13: gb_un.*
- 14: gb_vi.*
- 15: em_ba.*
- 16: em_fun.*
- 17: em_hum.*
- 18: em_in.*
- 19: em_mu.*
- 20: em_om.*
- 21: em_or.*
- 22: em_ov.*
- 23: em_pat.*
- 24: em_ph.*
- 25: em_pl.*
- 26: em_ro.*
- 27: em_sts.*
- 28: em_un.*
- 29: em_vi.*
- 30: em_htg_hum.*
- 31: em_htg_inv.*
- 32: em_htg_other.*
- 33: em_htg_mus.*
- 34: em_htg_pln.*
- 35: em_htg_rod.*
- 36: em_htg_mam.*
- 37: em_htg_vrt.*
- 38: em_sy.*
- 39: em_htgo_hum.*
- 40: em_htgo_mus.*
- 41: em_htgo_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	1557	100.0	1557	6	AX505141	AX505141 Sequence
2	1557	100.0	1560	8	AF467711	AF467711 Emericell
3	1557	100.0	1800	6	AX505252	AX505252 Sequence
4	1557	100.0	1833	6	AX505251	AX505251 Sequence
5	236.2	15.2	741	11	PM6B12G	AL686183 Penicilli
6	183	11.8	3751	8	AK110216	AK110216 Oryza sat
7	167.8	10.8	4310	8	AK1119806	AK119806 Oryza sat
8	61	3.9	7218	6	I66494	I66494 Sequence 14
9	59.6	3.8	241859	2	AC136406	AC136406 Rattus no
10	57.2	3.7	214690	2	AC083889	AC083889 Mus muscu
11	57.2	3.7	219559	10	AC087329	AC087329 Mus muscu
12	56.8	3.6	125020	9	AF429315	AF429315 Homo sapi
13	52.2	3.4	227533	2	AC094895	AC094895 Rattus no
14	52	3.3	178757	10	AC122509	AC122509 Mus muscu
15	51.8	3.3	197714	2	AC128967	AC128967 Rattus no
16	51.8	3.3	217985	2	BX649416	BX649416 Danio rer
17	51.6	3.3	264103	2	AC126651	AC126651 Rattus no
18	51	3.3	5452	6	AX646997	AX646997 Sequence
19	51	3.3	298166	2	AC087563	AC087563 Homo sapi
20	50.8	3.3	236054	2	AC114393	AC114393 Rattus no
21	50	3.2	110000	2	BX649525	Continuation (3 of
22	50	3.2	212125	5	BX324212	BX324212 Zebrafish
23	50	3.2	232885	2	AC106163	AC106163 Rattus no
24	50	3.2	233841	2	BX640576	BX640576 Danio rer
25	49.6	3.2	87120	2	AC012225	AC012225 Homo sapi
26	49.4	3.2	33058	2	AC100211	AC100211 Mus muscu
27	49.2	3.2	694	3	DMHSP70AB	Y14393 Drosophila
28	49.2	3.2	1751	10	MUSLRUNA	M34398 Mouse loric
29	49.2	3.2	1804	10	BC026781	BC026781 Mus muscu
30	49.2	3.2	1810	10	BC058223	BC058223 Mus muscu
31	49.2	3.2	1863	3	AY052029	AY052029 Drosophil
32	49.2	3.2	6530	6	AR076241	AR076241 Sequence
33	49.2	3.2	6530	10	MMU09189	U09189 Mus musculu
34	49.2	3.2	55528	3	AC105764	AC105764 Caenorhab
35	49.2	3.2	144152	3	AC007725	AC007725 Drosophil
36	49.2	3.2	182183	3	AC007889	AC007889 Drosophil
37	49.2	3.2	226782	3	AE003693	AE003693 Drosophil
38	49.2	3.2	255454	2	AC138676	AC138676 Mus muscu
39	49	3.1	189735	2	BX640537	BX640537 Danio rer
40	48.8	3.1	209887	2	AC079424	AC079424 Mus muscu
41	48.6	3.1	156506	2	AC068038	AC068038 Homo sapi
42	48.6	3.1	194864	10	AL691489	AL691489 Mouse DNA
43	48.6	3.1	223879	9	AC008735	AC008735 Homo sapi
44	48.4	3.1	149626	2	BX649341	BX649341 Danio rer
45	48.4	3.1	184211	2	BX072578	BX072578 Danio rer

ALIGNMENTS

RESULT 1
AX505141
LOCUS AX505141
DEFINITION Sequence 7 from Patent WO02057456.
ACCESSION AX505141
VERSION AX505141.1 GI:23386445
KEYWORDS
SOURCE synthetic construct
ORGANISM synthetic construct
artificial sequences.
REFERENCE 1
AUTHORS Cali,B.M., Madden,K.T., Milne,T.G., Zhang,L., Silva,J.C.,
Trueheart,J., Holtzman,D. and Sherman,A.
TITLE Regulators of fungal gene expression
JOURNAL Patent: WO 02057456-A 7 25-JUL-2002;

linear FAT 27-SEP-2002

Microbia (US)		Location/Qualifiers	
1. .1557		/organism="synthetic construct"	
		/mol_type="unassigned DNA"	
		/db_xref="taxon:32630"	
		/note="fungal gene"	
ORIGIN			
Query Match 100.0%; Score 1557; DB 6; Length 1557;			
Best Local Similarity 100.0%; Pred. No. 0;			
Matches 1557; Conservative 0; Mismatches 0; Indels 0; Gaps 0;			
QY	1	ATGGATCCTAGAAACCAATCCCTCTCGGCCTCCGTCACAGTCTGCCTCAAGATCGGCG	60
Db	1	ATGGATCCTAGAAACCAATCCCTCTCGGCCTCCGTCACAGTCTGCCTCAAGATCGGCG	60
QY	61	CCTCTTCCTTCTGCTCCCATCTCGAGCATGCCAATGCCTCAGTACACGATGCAGCCTCAG	120
Db	61	CCTCTTCCTTCTGCTCCCATCTCGAGCATGCCAATGCCTCAGTACACGATGCAGCCTCAG	120
QY	121	TACCCAGTCTCTCAGCCGCGACACCCCTGCCTCCTCTGCAACCCCATCATAGCCAGTCGCC	180
Db	121	TACCCAGTCTCTCAGCCGCGACACCCCTGCCTCCTCTGCAACCCCATCATAGCCAGTCGCC	180
QY	181	GCTCCTCACTCGTACATGGGGCAGCGCCGTTACCGGCCTGATCTGAACAGGTACCCCGCA	240
Db	181	GCTCCTCACTCGTACATGGGGCAGCGCCGTTACCGGCCTGATCTGAACAGGTACCCCGCA	240
QY	241	TCAAGTCACGATGTTTACGGGTCTTCTGCTGCGCGGATAATGCCCCACACTACCGTGGGC	300
Db	241	TCAAGTCACGATGTTTACGGGTCTTCTGCTGCGCGGATAATGCCCCACACTACCGTGGGC	300
QY	301	AGTTGCGCTCCGACATCTTTCCTTCTCATCCCAATCCGAGGCGCAGGCACAGCGCGAG	360
Db	301	AGTTGCGCTCCGACATCTTTCCTTCTCATCCCAATCCGAGGCGCAGGCACAGCGCGAG	360
QY	361	CAATCGCGCGCACTATCCTCCTCATAGCGTGTCCCGCGCGCTTCCAGCGCTCAGTCG	420
Db	361	CAATCGCGCGCACTATCCTCCTCATAGCGTGTCCCGCGCGCTTCCAGCGCTCAGTCG	420
QY	421	TACCGCGAGCCAAATTGCGCGCGCCTCCCGGACCGTCTCAACAATGGA	480
Db	421	TACCGCGAGCCAAATTGCGCGCGCCTCCCGGACCGTCTCAACAATGGA	480
QY	481	CTTCTCTCAGGAGCATTCAGTTATTTCGGACGGAAAGCCTCAAGTTGGACCCCGTTGCT	540
Db	481	CTTCTCTCAGGAGCATTCAGTTATTTCGGACGGAAAGCCTCAAGTTGGACCCCGTTGCT	540
QY	541	GCGAATGGTGTGTCGCGCGTATCCCGGGAAGGACTCCCGCGAAGGTTGTTGGTTCT	600
Db	541	GCGAATGGTGTGTCGCGCGTATCCCGGGAAGGACTCCCGCGAAGGTTGTTGGTTCT	600
QY	601	CAGGGCGACGCGGTATCCTTCCGAGTGTTCGGGACGCGCAACTCCGGTCACAAATGGT	660
Db	601	CAGGGCGACGCGGTATCCTTCCGAGTGTTCGGGACGCGCAACTCCGGTCACAAATGGT	660
QY	661	GTTAACGGCACCGGCAAGAACACTACTATCCCGGCAAGGATGCCGATGGAAAGTCCCT	720
Db	661	GTTAACGGCACCGGCAAGAACACTACTATCCCGGCAAGGATGCCGATGGAAAGTCCCT	720
QY	721	TGCCCCGAACGTAAACAAGACTTATCTTCATGCCAAGCATCTCAAGCGCCATCTGCTACGC	780
Db	721	TGCCCCGAACGTAAACAAGACTTATCTTCATGCCAAGCATCTCAAGCGCCATCTGCTACGC	780
QY	781	CACACTGGTGACCGCCCGTACATGTGTCTTTTGCAAAAGACACCTTCTCTCGCAGTGAT	840
Db	781	CACACTGGTGACCGCCCGTACATGTGTCTTTTGCAAAAGACACCTTCTCTCGCAGTGAT	840
QY	841	ATCCTGAAACGTCAATTTCCAAAAATGCTCAATCAGGCGTGGTAACCCACCGGAGCAACG	900
Db	841	ATCCTGAAACGTCAATTTCCAAAAATGCTCAATCAGGCGTGGTAACCCACCGGAGCAACG	900

QY	901	CACTTGTGCGACCCCAATGCGCATGTGAAGAGGTCCCAACAGCAGGCTCGGCGGAATCCT	960
Db	901	CACTTGTGCGACCCCAATGCGCATGTGAAGAGGTCCCAACAGCAGGCTCGGCGGAATCCT	960
QY	961	GTAATAACCTGTCCAGGATGAAGTCAGTAGTACCGTCCCGCTCCCAATGGCATCCCGGC	1020
Db	961	GTAATAACCTGTCCAGGATGAAGTCAGTAGTACCGTCCCGCTCCCAATGGCATCCCGGC	1020
QY	1021	ACGACTTACGGCAGGGAGCCGTCAACGGCAATGGACTAGTCCGGCCCGCAGGCTAC	1080
Db	1021	ACGACTTACGGCAGGGAGCCGTCAACGGCAATGGACTAGTCCGGCCCGCAGGCTAC	1080
QY	1081	GCGGATCACAGACTATGGGCTTCCCAATGTCCCAATGTCCATCCGTCAACGGGATGGCGGTGTCAG	1140
Db	1081	GCGGATCACAGACTATGGGCTTCCCAATGTCCCAATGTCCATCCGTCAACGGGATGGCGGTGTCAG	1140
QY	1141	CCTGAAGACCGGTTTCCCGCGCGCCGCGCATCAAGGAGCCCTTGGCCCAAGCTCCC	1200
Db	1141	CCTGAAGACCGGTTTCCCGCGCGCCGCGCATCAAGGAGCCCTTGGCCCAAGCTCCC	1200
QY	1201	AAGCAGAGCCCGTATCTCGTGACGCCGGGTGTGACCCCTTCTGGCCACCAGTTGAATATT	1260
Db	1201	AAGCAGAGCCCGTATCTCGTGACGCCGGGTGTGACCCCTTCTGGCCACCAGTTGAATATT	1260
QY	1261	GACCGAAACATCGAGCAGGTAAACAAACCGGTTGTTCAGACCCCAAGCGCCTGTGATG	1320
Db	1261	GACCGAAACATCGAGCAGGTAAACAAACCGGTTGTTCAGACCCCAAGCGCCTGTGATG	1320
QY	1321	CCAGGACATCCCGGCCACCCCGGTGAGCTTGACTGGACGTCTATGTTCCAAACCTCAAGCT	1380
Db	1321	CCAGGACATCCCGGCCACCCCGGTGAGCTTGACTGGACGTCTATGTTCCAAACCTCAAGCT	1380
QY	1381	CCCAGGGCTACATGTTCTCCCAGTCTATGCTGCTGAGTTCAGAGCCCAATCCACGCTCAT	1440
Db	1381	CCCAGGGCTACATGTTCTCCCAGTCTATGCTGCTGAGTTCAGAGCCCAATCCACGCTCAT	1440
QY	1441	GTCGAGACCGAGCGAAAGTATTACCCCAACCACTACCGTGGTCAAGAGAGTGGAAATGAAC	1500
Db	1441	GTCGAGACCGAGCGAAAGTATTACCCCAACCACTACCGTGGTCAAGAGAGTGGAAATGAAC	1500
QY	1501	GGTCTCTATCTGGCTTCGACTATGAGTGGCGACGCGACCGTTCAGCCCGCCAGACAA	1557
Db	1501	GGTCTCTATCTGGCTTCGACTATGAGTGGCGACGCGACCGTTCAGCCCGCCAGACAA	1557

RESULT 2

AF467711

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

source

gene

AF467711

Emericella nidulans

AF467711

AF467711.1

GI:27450538

Emericella nidulans

Emericella nidulans

Bukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes; Eurotiales; Trichocomaceae; Emericella.

1 (bases 1 to 1560)

Holtzman,D.A.

Enhanced Production of Secondary Metabolites in Filamentous Fungi by Ectopic Expression of Regulatory Proteins

Unpublished

2 (bases 1 to 1560)

Holtzman,D.A.

Direct Submission

Submitted (14-JAN-2002) Microbia, Inc., One Kendall Sq. Building 1400W, Cambridge, MA 02139, USA

Location/Qualifiers

1. .1560

/organism="Emericella nidulans"

/mol_type="mRNA"

/db_xref="taxon:162425"

1. .1560

/gene="rfec"

CDS
1. .1560
/gene="rfeC"
/note="putative transcription factor; contains zinc finger
and C2H2 domain; regulator of FLO11"
/codon_start=1
/product="RfeC"
/protein_id="AAO14631.1"
/db_xref="GI:27450539"
/translation="MDPRNHPSRPPSTSLPQGSAPLPSAPISSMPMPQYTNQPPQYPVS
QPHTLPLPQHSPAPHSYMGQPPYRDLNRPASSHDVYASSAAPIMHTTVGSL
PPTSFLSHPNPQAQASPHYPPHPSVLPASSAQSPQPIAPAPPRDRRADFNNG
LPSGAFSYSDGPKQWDVPAANGAAPYPGXDSPTQVVGSGRRGILPSVPGRATPVT
NGVNGTGKNTTIPAKDADGKFPNCPNKTYLHAKHLKRHLRLTRDPRYMCVLCKDTF
SRSDILKRHFQKCSIRGNPTGATHLSHPNAHVKRSQQQAAANPKVPQDEVSSVTPP
PNGIPGTTYEGAVNGGLAPARPGYADHQTWGFPMSSVNGMRGQPEDAFPGGRPHQ
GAPWPQAPKQSPYLVPQADPSGHQLNIDRNIETQVQPVVQDPKRPVMPGHPGHPGEL
DWTSMFQPAPEGYMESQSMPPGGQEPIDAHVETERKYVPTTTAGQESGMNGLYLASTM
SGDGTVPQPARQ"

ORIGIN

Query Match 100.0%; Score 1557; DB 8; Length 1560;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1557; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATGATCTAGAAACCATCCCTCTCGGCCTCCGTCTACCAAGTCTGCCTCAAGGATCGGCG 60
DB 1 ATGGATCTAGAAACCATCCCTCTCGGCCTCCGTCTACCAAGTCTGCCTCAAGGATCGGCG 60
QY 61 CCTCTCTCTCTGCTCCCATCTCGAGCATGCCAATGCCCTCAGTACACGATGCAGCCTCAG 120
DB 61 CCTCTCTCTCTGCTCCCATCTCGAGCATGCCAATGCCCTCAGTACACGATGCAGCCTCAG 120
QY 121 TACCCAGTCTCTCAGCCGCACACCCCTGCCTCCTCTGCAACCCCATCATAGCCAGTCGCC 180
DB 121 TACCCAGTCTCTCAGCCGCACACCCCTGCCTCCTCTGCAACCCCATCATAGCCAGTCGCC 180
QY 181 GCTCCTCACTCGTACATGCGGCGAGCCGCGGTACCGGCTGATCTGAACAGGTACCCCGCA 240
DB 181 GCTCCTCACTCGTACATGCGGCGAGCCGCGGTACCGGCTGATCTGAACAGGTACCCCGCA 240
QY 241 TCAAGTACAGATGTTTACGGGTCTTCTGCTGCGCCGATATGCCCCACACTACCGTGGGC 300
DB 241 TCAAGTACAGATGTTTACGGGTCTTCTGCTGCGCCGATATGCCCCACACTACCGTGGGC 300
QY 301 AGCTTGCTCCGACATCTTCTCTCATCCCAATCCGAGGCGCAGGCACAGGCGCAG 360
DB 301 AGCTTGCTCCGACATCTTCTCTCATCCCAATCCGAGGCGCAGGCACAGGCGCAG 360
QY 361 CAATCGCGCACTATCCTCCTCTCATAGCGTGTCTCCGCGCCGCTCCAGCGCTCAGTCG 420
DB 361 CAATCGCGCACTATCCTCCTCTCATAGCGTGTCTCCGCGCCGCTCCAGCGCTCAGTCG 420
QY 421 TACCCGCGCAATGCGCGCGGCGCTCCCGGGACCGTCTGCTGACTTCAACAATGGA 480
DB 421 TACCCGCGCAATGCGCGCGGCGCTCCCGGGACCGTCTGCTGACTTCAACAATGGA 480
QY 481 CTTCTTCAGGAGCATTCAGTTATTTCGACGGAAGCCTCAAGTTGGGACCCCGTTGCT 540
DB 481 CTTCTTCAGGAGCATTCAGTTATTTCGACGGAAGCCTCAAGTTGGGACCCCGTTGCT 540
QY 541 GCGAATGGTGTGCGCCCGTATCCCGGGAAGGACTCCCCCGAACCAGGTTGTTGTTCT 600
DB 541 GCGAATGGTGTGCGCCCGTATCCCGGGAAGGACTCCCCCGAACCAGGTTGTTGTTCT 600
QY 601 CAGGGCGACCGGGTATCCTTCCGAGTGTCCGGGACGCGCAACTCCGCTCACAATGGT 660
DB 601 CAGGGCGACCGGGTATCCTTCCGAGTGTCCGGGACGCGCAACTCCGCTCACAATGGT 660
QY 661 GTTAACGGCACCGGCAAGAACACTACTATCCCGGCCAAGGATGCCGATGGAAGTCCCT 720
DB 661 GTTAACGGCACCGGCAAGAACACTACTATCCCGGCCAAGGATGCCGATGGAAGTCCCT 720
QY 721 TGCCCGAACTGTAAACAAGACTTATCTTCATGCCAAGCATCTCAAGCGCCATCTGCTACGC 780

DB 721 TGCCCGAACTGTAAACAAGACTTATCTTCATGCCAAGCATCTCAAGCGCCATCTGCTACGC 780
QY 781 CACACTGGTGACCGCCGCTACATGTGTGTTCTTTGCAAGACACCTTCTCTCGCAGTGAT 840
DB 781 CACACTGGTGACCGCCGCTACATGTGTGTTCTTTGCAAGACACCTTCTCTCGCAGTGAT 840
QY 841 ATCCTGAAACGTCATTTCCAAAAAATGCTCAATCAGGCGTGTGTAACCCACCGGAGCAACG 900
DB 841 ATCCTGAAACGTCATTTCCAAAAAATGCTCAATCAGGCGTGTGTAACCCACCGGAGCAACG 900
QY 901 CACTTGTGCGACCCCAATGCGCATGTGAAGAGGTCCCAACAGCAGGCTGCGGCGAATCCT 960
DB 901 CACTTGTGCGACCCCAATGCGCATGTGAAGAGGTCCCAACAGCAGGCTGCGGCGAATCCT 960
QY 961 GTAAAAACCTGTCCAGGATGAAGTCAAGTACCGTCCCGCCTCCCAATGGCATCCCGGC 1020
DB 961 GTAAAAACCTGTCCAGGATGAAGTCAAGTACCGTCCCGCCTCCCAATGGCATCCCGGC 1020
QY 1021 ACGACTTACGGGAGGAGCCGTCAACGGCAATGGACTAGCTCCGGCCCGCCAGGGTAC 1080
DB 1021 ACGACTTACGGGAGGAGCCGTCAACGGCAATGGACTAGCTCCGGCCCGCCAGGGTAC 1080
QY 1081 GCGGATCACGACATATGGGCTTCCCAATGTATCCGTCAACGGGATGGCGCGTGGTCA 1140
DB 1081 GCGGATCACGACATATGGGCTTCCCAATGTATCCGTCAACGGGATGGCGCGTGGTCA 1140
QY 1141 CCTGAAGACGCTTCCCGGCGCGCGCGCATCAAGGAGCCCTTGGCCACAAGCTCCC 1200
DB 1141 CCTGAAGACGCTTCCCGGCGCGCGCGCATCAAGGAGCCCTTGGCCACAAGCTCCC 1200
QY 1201 AAGCAGAGCCCGTATCTCGTGACGCGGGTGTGACCCCTTCTGGCCACCAGTTGAATATT 1260
DB 1201 AAGCAGAGCCCGTATCTCGTGACGCGGGTGTGACCCCTTCTGGCCACCAGTTGAATATT 1260
QY 1261 GACCGAAACATCGAGCAGGTAAACAAACCGGTTGTTCAAGACCCCAAGCGCCTGTGATG 1320
DB 1261 GACCGAAACATCGAGCAGGTAAACAAACCGGTTGTTCAAGACCCCAAGCGCCTGTGATG 1320
QY 1321 CCAGGACATCCCGGCCACCCCGGTGAGCTTGACTGGACGTCTATGTTCCAACTCAAGCT 1380
DB 1321 CCAGGACATCCCGGCCACCCCGGTGAGCTTGACTGGACGTCTATGTTCCAACTCAAGCT 1380
QY 1381 CCCGAGGCTACATGTTCTCCAGTCTATGCTGCTGGTCAAGAGCCCATCCACGCTCAT 1440
DB 1381 CCCGAGGCTACATGTTCTCCAGTCTATGCTGCTGGTCAAGAGCCCATCCACGCTCAT 1440
QY 1441 GTCGAGACCGAGGAAAGTATTACCCCAACCACTACCGCTGCTCAAGAGAGTGGAATGAAC 1500
DB 1441 GTCGAGACCGAGGAAAGTATTACCCCAACCACTACCGCTGCTCAAGAGAGTGGAATGAAC 1500
QY 1501 GGTCTCTATCTGGCTTCGACTATGAGTGGCGACCGCAACCGTTACAGCCCGCCAGACAA 1557
DB 1501 GGTCTCTATCTGGCTTCGACTATGAGTGGCGACCGCAACCGTTACAGCCCGCCAGACAA 1557

RESULT 3
AX505252
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES

AX505252 1800 bp DNA linear PAT 27-SEP-2002
Sequence 118 from Patent WO02057456.

AX505252 GI:23386501

synthetic construct
synthetic construct
artificial sequences.

1
Cali, B.M., Madden, K.T., Milne, T.G., Zhang, L., Silva, J.C.,
Trueheart, J., Holtzman, D. and Sherman, A.
Regulators of fungal gene expression
Patent: WO 02057456-A 118 25-JUL-2002;
Microbia (US)

Location/Qualifiers

source	1. .1800 /organism="synthetic construct" /mol_type="unassigned DNA" /db_xref="taxon:32630" /note="fungal gene"									
ORIGIN										
Query Match 100.0%; Score 1557; DB 6; Length 1800; Best Local Similarity 100.0%; Pred. No. 0; Matches 1557; Conservative 0; Mismatches 0; Indels 0; Gaps 0;										
QY	1	ATGGATCCTAGAAACCATCCCTCTCGGGCTCCGTCTACCAGTCTGCCTCAAGGATCGGCG	60							
Db	244	ATGGATCCTAGAAACCATCCCTCTCGGGCTCCGTCTACCAGTCTGCCTCAAGGATCGGCG	303							
QY	61	CCTCTTCCTTCTGCTCCCATCTCGAGCATGCCAATGCCCTCAGTACACGATGCAGCCTCAG	120							
Db	304	CCTCTTCCTTCTGCTCCCATCTCGAGCATGCCAATGCCCTCAGTACACGATGCAGCCTCAG	363							
QY	121	TACCCAGTCTCTCAGCCGCACACCTGCTCCTCTGCAACCCCATCATAGCCAGTCGCC	180							
Db	364	TACCCAGTCTCTCAGCCGCACACCTGCTCCTCTGCAACCCCATCATAGCCAGTCGCC	423							
QY	181	GCTCCTCACTCGTACATGGGGCAGCCCGGTACCGGCCCTGATCTGAACAGGTACCCCGCA	240							
Db	424	GCTCCTCACTCGTACATGGGGCAGCCCGGTACCGGCCCTGATCTGAACAGGTACCCCGCA	483							
QY	241	TCAAGTCAGATGTTTACGGCTCTTCTGCTGCGCCGATATGCCCCACACTACCGTGGGC	300							
Db	484	TCAAGTCAGATGTTTACGGCTCTTCTGCTGCGCCGATATGCCCCACACTACCGTGGGC	543							
QY	301	AGTTTGCCTCCGACATCTTTCCCTTCTCATCCCAATCCGAGCGCGCAGGCAGGCGCAG	360							
Db	544	AGTTTGCCTCCGACATCTTTCCCTTCTCATCCCAATCCGAGCGCGCAGGCAGGCGCAG	603							
QY	361	CAATCGCGCACTATCCTCCTCATAGCGTGTCTCCGCCCGCTTCCAGCGCTCAGTCG	420							
Db	604	CAATCGCGCACTATCCTCCTCATAGCGTGTCTCCGCCCGCTTCCAGCGCTCAGTCG	663							
QY	421	TACCCGCGCAATTCGCGCGCGCTCCCGGACCGTCTGCTGACTTCAACAATGGA	480							
Db	664	TACCCGCGCAATTCGCGCGCGCTCCCGGACCGTCTGCTGACTTCAACAATGGA	723							
QY	481	CTTCCTTCAGGAGCATTCAGTTATTCCGACGGAAAGCCTCAAGTTGGACCCCGTTGCT	540							
Db	724	CTTCCTTCAGGAGCATTCAGTTATTCCGACGGAAAGCCTCAAGTTGGACCCCGTTGCT	783							
QY	541	GCGAATGGTGTGCGCGCGTATCCCGGAAGGACTCCCCCGGAACCCAGGTTGTTGTTCT	600							
Db	784	GCGAATGGTGTGCGCGCGTATCCCGGAAGGACTCCCCCGGAACCCAGGTTGTTGTTCT	843							
QY	601	CAGGGGCGACGGGTATCCTTCCGAGTGTTCGGGACGCGCAACTCCGGTCAAAAATGGT	660							
Db	844	CAGGGGCGACGGGTATCCTTCCGAGTGTTCGGGACGCGCAACTCCGGTCAAAAATGGT	903							
QY	661	GTTAACGGCACCGGCAAGAACACTACTATCCCGGCCAAGGATGCCGATGGAAGTTCCCT	720							
Db	904	GTTAACGGCACCGGCAAGAACACTACTATCCCGGCCAAGGATGCCGATGGAAGTTCCCT	963							
QY	721	TGCCCCGAACGTGTAAACAAGACTTATCTTCATGCCAAGCATCTCAAGCGCCATCTGCTACGC	780							
Db	964	TGCCCCGAACGTGTAAACAAGACTTATCTTCATGCCAAGCATCTCAAGCGCCATCTGCTACGC	1023							
QY	781	CACACTGGTGACCGCCCGGTACATGTGTGTTCTTTTGCAAAGACACCTTCTCTCGCAGTGAT	840							
Db	1024	CACACTGGTGACCGCCCGGTACATGTGTGTTCTTTTGCAAAGACACCTTCTCTCGCAGTGAT	1083							
QY	841	ATCCTGAAACGTCAATTTCCAAAAATGCTCAATCAGGCGTGGTAACCCCAACCGGAGCAACG	900							
Db	1084	ATCCTGAAACGTCAATTTCCAAAAATGCTCAATCAGGCGTGGTAACCCCAACCGGAGCAACG	1143							
QY	901	CACCTTGTGCAACCCCAATGCGCATGTGAAGAGGTCCCAACAGCAGGCTGCGGCGAATCCT	960							

Db	1144	CACTTGTGCAACCCCAATGCGCATGTGAAGAGGTCCCAACAGCAGGCTGCGGGAAATCCT	1203		
Qy	961	GTAAAACTGTCCAGATGAAGTCAGTAGTACCGTCCCGCCTCCCAATGGCATCCCGGC	1020		
Db	1204	GTAAAACTGTCCAGATGAAGTCAGTAGTACCGTCCCGCCTCCCAATGGCATCCCGGC	1263		
Qy	1021	ACGACTTACGGCGAGGAGCCGTCAACGGCAATGGAATAGTCTCGGCCCGGCCAGGTAC	1080		
Db	1264	ACGACTTACGGCGAGGAGCCGTCAACGGCAATGGAATAGTCTCGGCCCGGCCAGGTAC	1323		
Qy	1081	GCGGATCACCACTATGCGGCTTCCCAATGTCATCCGTCAACGGGATGGCCGTGTCAG	1140		
Db	1324	GCGGATCACCACTATGCGGCTTCCCAATGTCATCCGTCAACGGGATGGCCGTGTCAG	1383		
Qy	1141	CCTGAAGACGCGTTTCCCGCGCGCGCGCATCAAGGAGCCCTTGGCCACAAAGCTCCC	1200		
Db	1384	CCTGAAGACGCGTTTCCCGCGCGCGCGCATCAAGGAGCCCTTGGCCACAAAGCTCCC	1443		
Qy	1201	AAGCAGAGCCCGTATCTGTCAGCCGGGTGCTGACCCCTTCTGGCCACCAGTTGAATATT	1260		
Db	1444	AAGCAGAGCCCGTATCTGTCAGCCGGGTGCTGACCCCTTCTGGCCACCAGTTGAATATT	1503		
Qy	1261	GACCGAAACATCGAGAGGTAAACAAACCGGTTGTTCAAGACCCCAAGCGCCCTGTGATG	1320		
Db	1504	GACCGAAACATCGAGAGGTAAACAAACCGGTTGTTCAAGACCCCAAGCGCCCTGTGATG	1563		
Qy	1321	CCAGGACATCCCGGCACCCCGGTGAGCTTGACTTGAGCGTCTATGTTCCAACTCAAGCT	1380		
Db	1564	CCAGGACATCCCGGCACCCCGGTGAGCTTGACTTGAGCGTCTATGTTCCAACTCAAGCT	1623		
Qy	1381	CCCGAGGGCTACATGTTCTCCAGTCTATGCCCTGGTCAAGAGCCCATCCACGTCAT	1440		
Db	1624	CCCGAGGGCTACATGTTCTCCAGTCTATGCCCTGGTCAAGAGCCCATCCACGTCAT	1683		
Qy	1441	GTCGAGACCGGAGGAAGTATTACCCCAACCACTACCGTGGTCAAGAGAGTGGAAATGAAC	1500		
Db	1684	GTCGAGACCGGAGGAAGTATTACCCCAACCACTACCGTGGTCAAGAGAGTGGAAATGAAC	1743		
Qy	1501	GGTCTCTATCTGGCTTCGACTATGAGTGGCGACGGCACCGTTTCAGCCCGCCAGACAA	1557		
Db	1744	GGTCTCTATCTGGCTTCGACTATGAGTGGCGACGGCACCGTTTCAGCCCGCCAGACAA	1800		

RESULT 4

AX505251	LOCUS	AX505251	1833 bp	DNA	linear	PAT 27-SEP-2002
DEFINITION	Sequence 117 from Patent WO02057456.					
ACCESSION	AX505251					
VERSION	AX505251.1 GI:23386500					
KEYWORDS	.					
SOURCE	synthetic construct					
ORGANISM	synthetic construct					
	artificial sequences.					
REFERENCE	1					
AUTHORS	Cali,B.M., Madden,K.T., Milne,T.G., Zhang,L., Silva,J.C., Trueheart,J., Holtzman,D. and Sherman,A.					
TITLE	Regulators of fungal gene expression					
JOURNAL	Patent: WO 02057456-A 117 25-JUL-2002;					
	Microbia (US)					
FEATURES	Location/Qualifiers					
source	1. .1833					
	/organism="synthetic construct"					
	/mol_type="unassigned DNA"					
	/db_xref="taxon:32630"					
	/note="fungal gene"					
ORIGIN						

Query Match		100.0%; Score 1557; DB 6; Length 1833;			
Best Local Similarity		100.0%; Pred. No. 0;			
Matches 1557; Conservative		0; Mismatches 0; Indels 0; Gaps 0;			
Qy	1	ATGGATCCTAGAAACCATCCCTCTCGGCCCTCCGTCTACCAAGTCTGCCTCAAGGATCGGCG	60		

Db 277 ATGGATCCTAGAAACCATCCCTCTCGGCCTCCGTCTACCAAGTCTGCCCTCAAGGATCGGCG 336

QY 61 CCTCTTCCCTTCTGCTCCCATCTCGAGCATGCCAATGSCCTCAGTACACGATGAGCCTCAG 120

Db 337 CCTCTTCCCTTCTGCTCCCATCTCGAGCATGCCAATGSCCTCAGTACACGATGAGCCTCAG 396

QY 121 TACCAGTCTCTCAGCGGCACACCCCTGCCCTCTGTGCAACCCCATCATAGCCAGTGC GCC 180

Db 397 TACCAGTCTCTCAGCGGCACACCCCTGCCCTCTGTGCAACCCCATCATAGCCAGTGC GCC 456

QY 181 GCTCCTCACTCGTACATGGGGCAGCCGCGTACCGGCCCTGATCTGAACAGGTACCCCGCA 240

Db 457 GCTCCTCACTCGTACATGGGGCAGCCGCGTACCGGCCCTGATCTGAACAGGTACCCCGCA 516

QY 241 TCAAGTCAAGATGTTTACCGCTCTTCTGCTGCGCCGATAATGCCCCACACTACCGTGGC 300

Db 517 TCAAGTCAAGATGTTTACCGCTCTTCTGCTGCGCCGATAATGCCCCACACTACCGTGGC 576

QY 301 AGCTTGCCTCCGACATCTTTCCTTTCTCATCCCAATCCGAGGGCGAGGCAGCGCAG 360

Db 577 AGCTTGCCTCCGACATCTTTCCTTTCTCATCCCAATCCGAGGGCGAGGCAGCGCAG 636

QY 361 CAATCGCGCACTATCCTCCTCATAGCGTGCTCCCGCCGCTTCCAGCGCTCAGTCG 420

Db 637 CAATCGCGCACTATCCTCCTCATAGCGTGCTCCCGCCGCTTCCAGCGCTCAGTCG 696

QY 421 TACCCGCGAGCCAAATGCGCGCGCCCTCCCGGGACCGTCTGCTGACTTCAACAATGGA 480

Db 697 TACCCGCGAGCCAAATGCGCGCGCCCTCCCGGGACCGTCTGCTGACTTCAACAATGGA 756

QY 481 CTTCTCTCAGGAGCATTCAGTTATTTCGGACGGAAGCCTCAAGTTGGGACCCCGTTGCT 540

Db 757 CTTCTCTCAGGAGCATTCAGTTATTTCGGACGGAAGCCTCAAGTTGGGACCCCGTTGCT 816

QY 541 GCGAATGGTGTGCGCCGTATCCCGGGAGGACTCCCCCGAACCAGGTTGTTGGTTCT 600

Db 817 GCGAATGGTGTGCGCCGTATCCCGGGAGGACTCCCCCGAACCAGGTTGTTGGTTCT 876

QY 601 CAGGGCGACCGCGTATCCTTCCGAGTGTCCGGGACGCGCAACTCCGGTCACAAATGGT 660

Db 877 CAGGGCGACCGCGTATCCTTCCGAGTGTCCGGGACGCGCAACTCCGGTCACAAATGGT 936

QY 661 GTTAACGGCACCGGCAAGAACACTACTATCCCGGCCAAGGATGCCGATGGAAAGTTCCCT 720

Db 937 GTTAACGGCACCGGCAAGAACACTACTATCCCGGCCAAGGATGCCGATGGAAAGTTCCCT 996

QY 721 TGCCCGAACTGTAACAAGACTTATCTTCTATGTCCTCAAGCATCTCAAGGCCATCTGTACGC 780

Db 997 TGCCCGAACTGTAACAAGACTTATCTTCTATGTCCTCAAGCATCTCAAGGCCATCTGTACGC 1056

QY 781 CACACTGGTGACCGCCCGTACATGTGTGTTCTTTGCAAGACACCTTCTCTCGCAGTGAT 840

Db 1057 CACACTGGTGACCGCCCGTACATGTGTGTTCTTTGCAAGACACCTTCTCTCGCAGTGAT 1116

QY 841 ATCCTGAACCGTCATTTCCAAAATGCTCAATCAGGCGTGGTAAACCCACCGGAGCAACG 900

Db 1117 ATCCTGAACCGTCATTTCCAAAATGCTCAATCAGGCGTGGTAAACCCACCGGAGCAACG 1176

QY 901 CACTTGTGCAACCCCAATGCGCATGTGAAGAGGTCCCAACAGCAGGTGCGGCGAATCCT 960

Db 1177 CACTTGTGCAACCCCAATGCGCATGTGAAGAGGTCCCAACAGCAGGTGCGGCGAATCCT 1236

QY 961 GTAAAACCTGTCCAGGATGAAGTCAGTACCGTCCCGCTCCCAATGGCATCCCGGC 1020

Db 1237 GTAAAACCTGTCCAGGATGAAGTCAGTACCGTCCCGCTCCCAATGGCATCCCGGC 1296

QY 1021 ACGACTTACGGGAGGGAGCCGTCAACGGCAATGGACTAGCTCCGSCCGGCCAGGGTAC 1080

Db 1297 ACGACTTACGGGAGGGAGCCGTCAACGGCAATGGACTAGCTCCGSCCGGCCAGGGTAC 1356

QY 1081 GCGGATCACCACTATGGGCTTCCCAATGTCCGTCAACGGGATGGGCCGTGGTCAG 1140

Db 1357 GCGGATCACCACTATGGGCTTCCCAATGTCCGTCAACGGGATGGGCCGTGGTCAG 1416

QY 1141 CCTGAAGACGCGTTTCCCGGCGCGCGCATCAAGGAGCCCTTGGCCACAAGCTCCC 1200

Db 1417 CCTGAAGACGCGTTTCCCGGCGCGCGCATCAAGGAGCCCTTGGCCACAAGCTCCC 1476

QY 1201 AAGCAGAGCCCGTATCTCGTGCAGCCGGGTGCTGACCCCTTCTGGCCACCAGTTGAATATT 1260

Db 1477 AAGCAGAGCCCGTATCTCGTGCAGCCGGGTGCTGACCCCTTCTGGCCACCAGTTGAATATT 1536

QY 1261 GACCGAAACATCGAGCAGGTAAACAACCGGTTGTTCAAGACCCCAAGCGCCTGTGATG 1320

Db 1537 GACCGAAACATCGAGCAGGTAAACAACCGGTTGTTCAAGACCCCAAGCGCCTGTGATG 1596

QY 1321 CCAGGACATCCCGGCCACCCCGGTGAGCTTGACTGGACGTCTATGTTCCAAACCTCAAGCT 1380

Db 1597 CCAGGACATCCCGGCCACCCCGGTGAGCTTGACTGGACGTCTATGTTCCAAACCTCAAGCT 1656

QY 1381 CCCGAGGGCTACATGTTTCTCCAGTCTATGCCTGGTGGTCAAGAGCCCATCCACGCTCAT 1440

Db 1657 CCCGAGGGCTACATGTTTCTCCAGTCTATGCCTGGTGGTCAAGAGCCCATCCACGCTCAT 1716

QY 1441 GTCGAGACCGAGCGAAAGTATTACCCCACTACTACCGTGGTCAAGAGAGTGAATGAAC 1500

Db 1717 GTCGAGACCGAGCGAAAGTATTACCCCACTACTACCGTGGTCAAGAGAGTGAATGAAC 1776

QY 1501 GGTCTCTATCTGGCTTCGACTATGAGTGGCGACGGCACCGTTCAGCCCGCCAGACAA 1557

Db 1777 GGTCTCTATCTGGCTTCGACTATGAGTGGCGACGGCACCGTTCAGCCCGCCAGACAA 1833

RESULT 5

PM6B12G

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Penicillium marneffe

Penicillium marneffe

Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;

Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Penicillium.

1

AUTHORS

Yuen,K.Y., Pascal,G., Wong,S.S., Glaser,P., Woo,P.C., Kunst,F.,

Cai,J.J., Cheung,E.Y., Medigue,C. and Danchin,A.

Exploring the Penicillium marneffe genome

Arch. Microbiol. 179 (5), 339-353 (2003)

22595073

12640520

2 (bases 1 to 741)

Danchin,A. and Pascal,G.

Direct Submission

Submitted (08-MAR-2002) Danchin A., HKU-Pasteur Research Centre,

Dexter HC Man Building 8, Sassoon Road, Pokfulam, Hong Kong

Location/Qualifiers

1. .741

/organism="Penicillium marneffe"

/mol_type="genomic DNA"

/db_xref="taxon:37727"

/clone="pm6b12.g"

ORIGIN

Query Match

Best Local Similarity

Matches

457; Conservative

0; Mismatches

180; Indels

74; Gaps

4;

QY 601 CAGGGCGACGCGGTATCCTTCGAGTGTTCGGGACGCGCAACTCCGGTCACAAATGGT 660

Db 2 CAGGGACGACGAGGCAATTCTCCAGTGTTCGGGACGCG--CTGCAGTCCCCAACGGT 58

QY 661 GTTAACGGCACCGGCAAGAACACTACTATCCCGGCCAAGGATGCCGATGGAAAGTTCCCT 720

Db 59 GCCATGAACCGCTAGAACCAACCGATTCCGGCCCAAGGATGCCGACGGCAAGTTCCCT 118

PM6B12G

Penicillium marneffe STS, clone pm6b12.g, sequence tagged site.

AL686183

AL686183.1 GI:19337912

STS.

Penicillium marneffe

Penicillium marneffe

Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;

Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Penicillium.

1

AUTHORS

Yuen,K.Y., Pascal,G., Wong,S.S., Glaser,P., Woo,P.C., Kunst,F.,

Cai,J.J., Cheung,E.Y., Medigue,C. and Danchin,A.

Exploring the Penicillium marneffe genome

Arch. Microbiol. 179 (5), 339-353 (2003)

22595073

12640520

2 (bases 1 to 741)

Danchin,A. and Pascal,G.

Direct Submission

Submitted (08-MAR-2002) Danchin A., HKU-Pasteur Research Centre,

Dexter HC Man Building 8, Sassoon Road, Pokfulam, Hong Kong

Location/Qualifiers

1. .741

/organism="Penicillium marneffe"

/mol_type="genomic DNA"

/db_xref="taxon:37727"

/clone="pm6b12.g"

Query Match

Best Local Similarity

Matches

457; Conservative

0; Mismatches

180; Indels

74; Gaps

4;

QY 601 CAGGGCGACGCGGTATCCTTCGAGTGTTCGGGACGCGCAACTCCGGTCACAAATGGT 660

Db 2 CAGGGACGACGAGGCAATTCTCCAGTGTTCGGGACGCG--CTGCAGTCCCCAACGGT 58

QY 661 GTTAACGGCACCGGCAAGAACACTACTATCCCGGCCAAGGATGCCGATGGAAAGTTCCCT 720

Db 59 GCCATGAACCGCTAGAACCAACCGATTCCGGCCCAAGGATGCCGACGGCAAGTTCCCT 118

QY	721	TGCCCCGAAC	TGTAACAAGACTTATCTTCATGCCAAGCATCTCAAGCGGCATCTGCTACGC	780
Db	119	TGCCCTCAT	TGTGACCAAGACGTACCTCCACGGAAGCATCTTAAACGCCATCTTTTGAGA	178
QY	781	C-----	-----ACA	784
Db	179	CGTAAGTTG	TCCACTCAATTACCTTTTCCATCTATTAAACCAATCAAAACGCAGACA	238
QY	785	CTGGTGAC	CGCGGTACATGTGTGTTCTTTGCAAAAGACACCTTCTCTCGCAGTGATATCC	844
Db	239	CCGGTGAT	CGCCCTACATGTGTGTTCTCTGCAAAAGATACATTCTCCGGAAGCGATATTT	298
QY	845	TGAAACGT	CATTTCGAAAAAATGCTCAATCAGCGGTGGTAAACCCACCGGAGCAACGCCACT	904
Db	299	TGAAACGA	CACCTTCGAAAAAGTGCTCGTTACGCGCGGTAAACCCACAGGAGCGACCATC	358
QY	905	TGTCGCAC	CCCCAATGTCGCATGTGAAGAGGTTCCCAACAGCAGCGCTGCGGCGAATCCTGTAA	964
Db	359	TCTCGCAT	CCTCAGGCGCACCTGAAACGGTC-----TCAGGCGCGAGCTAATGCTGCAA	412
QY	965	AACCTGT	CCAGATGAAGTCAGTAGTACCGTCCCGGCCTCCCAATGGCATCCCGGGCACGA	1024
Db	413	AATCCGTT	CAGAAAGTCAGTAACTCCATTCGCGCCTTCCAATGGTATTGCGGGTACGA	472
QY	1025	CTTACGG	CGGAGGAGCCGTCACACGGCAATGAGTACTAGTCCGGC-----CCGGCCAG	1075
Db	473	CTTATGG	CGACGCGCACGCTGAATGGCCATGGCCACACGCTGGCTCCGGTGGCCGACCTG	532
QY	1076	GGTACG	CGGATCACACGACTATGGGCTTCCCAATGTCCATCCGTCAACGGGATGGCCGTG	1135
Db	533	GTCTCAC	GGAGCAGCCCTCTGGGATATTCAATGTCTCTGCCCCAGCGGTGAACAAATG	592
QY	1136	GTCAGC	CTGAACGCGTTTCCCGGCGCGCGCATCAAGGAGCGCCCTTGGCCACAAG	1195
Db	593	GTCACG	TGTGGACGCGTTCAACCCCTGCCAGGCAGATCACAGAGCTTCTGGATGGCTG	652
QY	1196	CTCCCA	AGCAGAGCCCGTATCTCGTGCAGCCGGTGCTGACCCCTTCTGGCC	1246
Db	653	CTCAGA	AGCAGAGCCCGTACCTTGTNCACTCTTCTGGTCTGTGTAAGGTC	703

RESULT	6
AK110216	
LOCUS	AK110216 3751 bp mRNA linear PLN 24-JUL-2003
DEFINITION	Oryza sativa (japonica cultivar-group) cDNA clone:002-162-D06, full insert sequence.
ACCESSION	AK110216
VERSION	AK110216.1 GI:32995425
KEYWORDS	FLI_CDNA; oligo capping.
SOURCE	Oryza sativa (japonica cultivar-group)
ORGANISM	Oryza sativa (japonica cultivar-group) Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae; Oryza.
REFERENCE	1
AUTHORS	The Rice Full-length cDNA Consortium, National Institute of Agrobiological Sciences Rice Full-length cDNA Project Team: Kikuchi, S., Nagata, T., Hayatsu, N., Doi, K., Kishimoto, N., Yazaki, J., Ishikawa, M., Yamada, H., Ooka, H., Hotta, I., Kojima, K., Namiki, T., Ohneda, E., Yahagi, W., Suzuki, K., Li, C., Ohtsuki, K., Shishiki, T., Foundation of Advancement of International Science Genome Sequencing & Analysis Group: Otomo, Y., Murakami, K., Iida, Y., Sugano, S., Fujimura, T., Masuda, H., Kobayashi, M., Xie, Q., Lu, M., Kurosaki, T., Kodama, T., Masuda, H., Kobayashi, M., Xie, Q., Lu, M., Narikawa, R., Sugiyama, A., Mizuno, K., Yokomizo, S., Niikura, J., Ikeda, R., Ishibiki, J., Kawamata, M., Yoshimura, A., Miura, J., Kusumegi, T., Oka, M., Ryu, R., Ueda, M., Matsubara, K., RIKEN: Kawai, J., Carninci, P., Adachi, J., Aizawa, K., Arakawa, T., Fukuda, S., Hara, A., Hashidume, W., Hayatsu, N., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kondo, S., Konno, H., Miyazaki, A., Osato, N., Ota, Y., Saito, R., Sasaki, D., Sato, K., Shibata, K., Shinagawa, A., Shiraki, T., Yoshino, M. and Hayashizaki, Y.
TITLE	Collection, mapping, and annotation of over 28,000 cDNA clones from

JOURNAL	japonica rice
MEDLINE	Science 301 (5631), 376-379 (2003)
PUBMED	22752273
REFERENCE	12869764
AUTHORS	2 (bases 1 to 3751) Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Doi, K., Fujimura, T., Fukuda, S., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayashizaki, Y., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hori, F., Hotta, I., Iida, J., Iida, Y., Ikeda, R., Imamura, K., Imotani, K., Ishibiki, J., Ishii, Y., Ishikawa, M., Itoh, M., Kagawa, I., Kanagawa, S., Katoh, H., Kawagashira, N., Kawai, J., Kawamata, M., Kikuchi, S., Kishikawa-Hirozane, T., Kishimoto, N., Kobayashi, M., Kodama, T., Kojima, K., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Kurosaki, T., Kusumegi, T., Li, C., Lu, M., Masuda, H., Matsubara, K., Murata, M., Nagata, T., Niikura, J., Miyazaki, A., Mizuno, K., Murakami, K., Niikura, J., Nishi, K., Nomura, K., Namiki, T., Narikawa, R., Niikura, J., Ohneda, E., Ohtsuki, K., Oka, M., Ooka, H., Numasaki, R., Ohneda, E., Ohno, M., Ohtsuki, K., Sakai, C., Sakai, K., Osato, N., Ota, Y., Otomo, Y., Ryu, R., Saitoh, H., Sakai, C., Sasaki, K., Sakazume, N., Sano, H., Sasaki, D., Sato, K., Shibata, K., Shibata, K., Sugiyama, A., Suzuki, K., Suzuki, Y., Tagami, M., Tagami-Takeda, Y., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Tanaka, T., Tomaru, A., Toya, T., Tsunoda, Y., Ueda, M., Waki, K., Xie, Q., Yahagi, W., Yamada, H., Yamamoto, M., Yasunishi, A., Yazaki, J., Yokomizo, S. and Yoshimura, A.
TITLE	Direct Submission
JOURNAL	Submitted (27-AUG-2002) Shoshi Kikuchi, National Institute of Agrobiological Sciences, Department of Molecular Genetics, Head of Laboratory of Gene Expression; 2-1-2 Kannondai, Tsukuba, Ibaraki 305-8602, Japan (E-mail:skikuchi@nias.affrc.go.jp, Tel:81-29-838-7007, Fax:81-29-838-7007)
COMMENT	This clone is one of the 28K full-length cDNA clones from japonica rice. URL : http://cdna01.dna.affrc.go.jp/cdna/ NIAS Rice Full-length cDNA Project Team: Kikuchi, S., Sato, K., Nagata, T., Kawagashira, N., Doi, K., Kishimoto, N., Yazaki, J., Ishikawa, M., Yamada, H., Ooka, H., Hotta, I., Kojima, K., Namiki, T., Ohneda, E., Yahagi, W., Suzuki, K., Li, C., Ohtsuki, K., Shishiki, T. and Yamamoto, M. FAIS Genome Sequencing & Analysis Group: Otomo, Y., Iida, Y., Fujimura, T., Ikeda, R., Ishibiki, J., Kawamata, M., Kobayashi, M., Kodama, T., Kurosaki, T., Kusumegi, T., Lu, M., Masuda, H., Miura, J., Mizuno, K., Narikawa, R., Niikura, J., Oka, M., Ryu, R., Sugano, S., Sugiyama, A., Suzuki, Y., Tsunoda, Y., Ueda, M., Xie, Q., Yokomizo, S., Yoshimura, A., Matsubara, K. and Murakami, K. Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken: Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda, S., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hori, F., Iida, J., Imamura, K., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kanagawa, S., Katoh, H., Kawai, J., Kishikawa-Hirozane, T., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numasaki, R., Ohno, M., Osato, N., Ota, Y., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Sato, K., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagami-Takeda, Y., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Tanaka, T., Tomaru, A., Toya, T., Waki, K., Yasunishi, A. and Hayashizaki, Y.
FEATURES	Location/Qualifiers 1..3751 /organism="Oryza sativa (japonica cultivar-group)" /mol_type="mRNA" /cultivar="Nipponbare" /db_xref="taxon:39947" /clone="002-162-D06"
source	
ORIGIN	
Query Match	11.8%; Score 183; DB 8; Length 3751;
Best Local Similarity	61.0%; Pred. No. 2.3e-35;
Matches 335; Conservative	0; Mismatches 205; Indels 9; Gaps 2;

Weinstock,G. and Gibbs,R.A.
Direct Submission
Unpublished
2 (bases 1 to 241859)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (01-NOV-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 241859)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (23-NOV-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Nov 23, 2002 this sequence version replaced gi:24462281.
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
in the feature table below represents a scaffold in the Atlas
assembly (a 'contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented, and separated
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
shotgun sequence only contigs will be indicated in the feature
table.
----- Genome Center -----
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information -----
Center project name: KCQC
Center clone name: CH230-522A18
----- Summary Statistics -----
Assembly program: Phrap; version 0.990329
Consensus quality: 187090 bases at least Q40
Consensus quality: 189991 bases at least Q30
Consensus quality: 191799 bases at least Q20
Estimated insert size: 211470; sum-of-contigs estimation
Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 5 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 64842: contig of 64842 bp in length
* 64843 64942: gap of unknown length
* 64943 70527: contig of 5585 bp in length
* 70528 70627: gap of unknown length
* 70628 237692: contig of 167065 bp in length
* 237693 237792: gap of unknown length
* 237793 238825: contig of 1033 bp in length
* 238826 238925: gap of unknown length
* 238926 241859: contig of 2934 bp in length.
FEATURES
source
1. .241859
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-522A18"
2456. .3351
/note="clone boundary
clone end:Sp6
site:MboI
misc_feature
end_sequence:BZ285495"
complement(52603..53490)
/note="clone_boundary
clone end:T7
site:MboI
end sequence:BZ285493"
102059. .103179
/note="wgs_end_extension
clone end:T7"
113425. .117805
/note="wgs_end_extension
clone end:T7"
ORIGIN
Query Match 3.8%; Score 59.6; DB 2; Length 241859;
Best Local Similarity 45.9%; Pred. No. 0.00036;
Matches 203; Conservative 0; Mismatches 239; Indels 0; Gaps 0;
QY 16 CATCCCTCTGGCCTCCGTCTACCAAGTCTGCCTCAAGGATCGGCGCCTCTTCCCTTCTGCT 75
Db 87310 CCTCTGCCTCTGCCTCTGCCTCTGCCTCTGCCTCTGCCTCTGCCTCTGCCTCTGCCT 87251
QY 76 CCCATCTCGAGCATGCCAATGCCTCAGTACACGATGCAGCCCTAGTACCCAGTCTCTCAG 135
Db 87250 CTGCCTCTGCCTCTGCCTCTGCCTCTGCCTCTGCCTCTGCCTCTGCCTCTGCCTCTG 87191
QY 136 CCGCACACCCCTGCCTCCTCTGGAACCCCATCATAGCCAGTGCCTGCTCCTCACTCGTAC 195
Db 87190 CCTCTGCCTCTGCCTCTGCCTCTGCCTCTGCCTCTGCCTCTGCCTCTGCCTCTGCCT 87131
QY 196 ATGGGCGAGCGCCGTACCGGCTGATCTGAACAGGTACCCCGCATCAAGTCACGATGTT 255
Db 87130 CCTCTGCCTCTGCCTCTGCCTCTGCCTCTGCCTCTGCCTCTGCCTCTGCCTCTGCCT 87071
QY 256 TACGGCTCTTCTGTGCGCCGATGAATGCCCCACACACTACCGTGGGAGCTTGCCTCCGACA 315
Db 87070 CCTCTGCCTCTGCCTCTGCCTCTGCCTCTGCCTCTGCCTCTGCCTCTGCCTCTGCCT 87011
QY 316 TCTTCTCTTCTCATCCCAATCCGCGAGGCGCAGGCAGGCAGCAATGCCGCACTAT 375
Db 87010 CCTCTGCCTCTGCCTCTGCCTCTGCCTCTGCCTCTGCCTCTGCCTCTGCCTCTGCCT 86951
QY 376 CCTCTCTCTCATAGCGTCTCCCGCCCGCTTCCAGCGCTCAGTCGTCACCCGAGCCAAAT 435
Db 86950 CCTCTGCCCTCTGCCCTGCCCTGCCCTGCCCTGCCCTGCCCTGCCCTGCCCTCTG 86891
QY 436 GCGCGCGGCGCTCCCGGGGACC 457
Db 86890 CCTCTGCCCTGACCCCTGCCCTGCCCTGCCCTGCCCTGCCCTGCCCTGCCCTCTG 86869
RESULT 10
AC083889/c
LOCUS
DEFINITION
Mus musculus chromosome 5 clone RP23-271A20 strain C57BL6/J,
WORKING DRAFT SEQUENCE, 29 unordered pieces.
AC083889
AC083889.1 GI:10645305
HTG; HTGS PHASE1; HTGS_DRAFT.
KEYWORDS
SOURCE
Mus musculus (house mouse)
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1 (bases 1 to 214690)
Beckstrom-Sternberg,S.M., Benjamin,B., Blakesley,R.W.,
Bouffard,G.G., Dietrich,N.L., Gupta,J., Ho,S.-L., Idol,J.,
Lee-Lin,S.-Q., Legaspi,R., Lim,M., Maduro,Q.L., Maduro,V.B.,
Mastrian,S.D., McCloskey,J.C., McDowell,J., Pearson,R., Snyder,B.,
Stantripop,S., Thomas,J.W., Thomas,P.J., Tjongson,E.E.,
Touchman,J.W., Tran,J.T., Vogt,J.L., Walker,M.A., Wetherby,K.D. and
Green,E.D.
NISC Mouse Sequencing Initiative
Unpublished
TITLE
JOURNAL

misc_feature
misc_feature
misc_feature
ORIGIN
Query Match 3.8%; Score 59.6; DB 2; Length 241859;
Best Local Similarity 45.9%; Pred. No. 0.00036;
Matches 203; Conservative 0; Mismatches 239; Indels 0; Gaps 0;
QY 16 CATCCCTCTGGCCTCCGTCTACCAAGTCTGCCTCAAGGATCGGCGCCTCTTCCCTTCTGCT 75
Db 87310 CCTCTGCCTCTGCCTCTGCCTCTGCCTCTGCCTCTGCCTCTGCCTCTGCCTCTGCCT 87251
QY 76 CCCATCTCGAGCATGCCAATGCCTCAGTACACGATGCAGCCCTAGTACCCAGTCTCTCAG 135
Db 87250 CTGCCTCTGCCTCTGCCTCTGCCTCTGCCTCTGCCTCTGCCTCTGCCTCTGCCTCTG 87191
QY 136 CCGCACACCCCTGCCTCCTCTGGAACCCCATCATAGCCAGTGCCTGCTCCTCACTCGTAC 195
Db 87190 CCTCTGCCTCTGCCTCTGCCTCTGCCTCTGCCTCTGCCTCTGCCTCTGCCTCTGCCT 87131
QY 196 ATGGGCGAGCGCCGTACCGGCTGATCTGAACAGGTACCCCGCATCAAGTCACGATGTT 255
Db 87130 CCTCTGCCTCTGCCTCTGCCTCTGCCTCTGCCTCTGCCTCTGCCTCTGCCTCTGCCT 87071
QY 256 TACGGCTCTTCTGTGCGCCGATGAATGCCCCACACACTACCGTGGGAGCTTGCCTCCGACA 315
Db 87070 CCTCTGCCTCTGCCTCTGCCTCTGCCTCTGCCTCTGCCTCTGCCTCTGCCTCTGCCT 87011
QY 316 TCTTCTCTTCTCATCCCAATCCGCGAGGCGCAGGCAGGCAGCAATGCCGCACTAT 375
Db 87010 CCTCTGCCTCTGCCTCTGCCTCTGCCTCTGCCTCTGCCTCTGCCTCTGCCTCTGCCT 86951
QY 376 CCTCTCTCTCATAGCGTCTCCCGCCCGCTTCCAGCGCTCAGTCGTCACCCGAGCCAAAT 435
Db 86950 CCTCTGCCCTCTGCCCTGCCCTGCCCTGCCCTGCCCTGCCCTGCCCTGCCCTCTG 86891
QY 436 GCGCGCGGCGCTCCCGGGGACC 457
Db 86890 CCTCTGCCCTGACCCCTGCCCTGCCCTGCCCTGCCCTGCCCTGCCCTGCCCTCTG 86869
RESULT 10
AC083889/c
LOCUS
DEFINITION
Mus musculus chromosome 5 clone RP23-271A20 strain C57BL6/J,
WORKING DRAFT SEQUENCE, 29 unordered pieces.
AC083889
AC083889.1 GI:10645305
HTG; HTGS PHASE1; HTGS_DRAFT.
KEYWORDS
SOURCE
Mus musculus (house mouse)
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1 (bases 1 to 214690)
Beckstrom-Sternberg,S.M., Benjamin,B., Blakesley,R.W.,
Bouffard,G.G., Dietrich,N.L., Gupta,J., Ho,S.-L., Idol,J.,
Lee-Lin,S.-Q., Legaspi,R., Lim,M., Maduro,Q.L., Maduro,V.B.,
Mastrian,S.D., McCloskey,J.C., McDowell,J., Pearson,R., Snyder,B.,
Stantripop,S., Thomas,J.W., Thomas,P.J., Tjongson,E.E.,
Touchman,J.W., Tran,J.T., Vogt,J.L., Walker,M.A., Wetherby,K.D. and
Green,E.D.
NISC Mouse Sequencing Initiative
Unpublished
TITLE
JOURNAL


```
misc_feature      /note="assembly_fragment"
127280..139420
misc_feature      /note="assembly_fragment"
139521..148942
misc_feature      /note="assembly_fragment"
149043..162179
misc_feature      /note="assembly_fragment"
162280..177494
misc_feature      /note="assembly_fragment"
177595..190534
misc_feature      /note="assembly_fragment"
190635..214690
misc_feature      /note="assembly_fragment"

ORIGIN
Query Match      3.7%; Score 57.2; DB 2; Length 214690;
Best Local Similarity 45.7%; Pred. No. 0.0015;
Matches 200; Conservative 0; Mismatches 238; Indels 0; Gaps 0;

QY 16 CATCCCTCTCGGCTCCGTCTACCAAGTCTGCCTCAAGGATCGGCGCCTCTTCTGCT 75
Db 194143 CCTCTGCCTCTGCCTCTGCCTCTGCCTCTGCCTCTGCCTCTGCCTCTGCTG 194084

QY 76 CCCATCTCGAGCATGCCAATGCCTCAGTACAGCATGCAGCTCAGTACCCAGTCTCTCAG 135
Db 194083 CCTCCTCTGCCTCTGCCTCTGCCTCTGCCTCTGCCTCTGCCTCTGCCTCTCT 194024

QY 136 CCGCACACCCCTGCTCCTCTGCAACCCCATCATAGCCAGTCGCCCGCTCCTCACTCGTAC 195
Db 194023 CTGCCCTCTGCTGCTCCTCTGCCTCCTCTGCTCCTCTGCTCCTCTGCTCCTCTG 193964

QY 196 ATGGGGCAGCGCGGTACCGGCTGATCTGAACAGGTACCCCGCATCAAGTCACGATGTT 255
Db 193963 CTGCCCTCTGCCTCTGCCTCTGCCTCTGCCTCTGCCTCTGCCTCTGCCTCTGCT 193904

QY 256 TAGCGGTCTCTGCTGCGCGGATATGCCACACTACCGTGGCAGCTTGCTCCGACA 315
Db 193903 CCTCTGCCCTCCTGCTCCTCTGCCTCCTGCTCCTGCTCCTGCTCCTGCTCCTG 193844

QY 316 TCTTTCCTTTTCATCCCAATCCGAGGCGGAGGCACAGGCGCAATCGCCGCACTAT 375
Db 193843 CCTCCTCTGCCTCTGCTGCTCCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 193784

QY 376 CCTCCTCCTCATAGCGTGTCCCGCCGCTTCCAGCGCTCAGTCGTACCGCAGCCAATT 435
Db 193783 CCTCTGCCCTCCTGCTGCTCCTCTGCTGCTCCTGCTCCTGCTGCTGCTGCTGCT 193724

QY 436 GCGCGGCGGCTCCCGG 453
Db 193723 CCTCCTCTGCCTCTGCTG 193706
```

```
RESULT 11
AC087329
LOCUS      AC087329      219559 bp      DNA      linear      ROD 23-OCT-2002
DEFINITION Mus Musculus Strain C57BL6/J chromosome 5 BAC, RP23-383N15,
            Complete Sequence, complete sequence.
ACCESSION  AC087329
VERSION     AC087329.10 GI:24270683
KEYWORDS   HTG.
SOURCE     Mus musculus (house mouse)
ORGANISM   Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE  1 (bases 1 to 219559)
AUTHORS    Montgomery,K.T., Grills,G., Han,J., Chiu,D., Decker,J., Fusina,M.,
            Goltz,J., Haider,A., Hall,L., Ioshikhes,I.P., Lee,E., Long,J.,
            Perera,A., Shim,C., Thomas,E. and Kucherlapati,R.
            High Throughput Mouse Sequencing
            Unpublished
TITLE      Unpublished
JOURNAL    2 (bases 1 to 219559)
REFERENCE  2 (bases 1 to 219559)
AUTHORS    Montgomery,K.T., Grills,G., Han,J., Chiu,D., Decker,J., Fusina,M.,
            Goltz,J., Haider,A., Hall,L., Ioshikhes,I.P., Lee,E., Long,J.,
```

```
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
-----Genome Center:
Center: Harvard Partners Genome Center
Center Code: HPGC
Web site: http://www.hpcgg.org/Sequence/mouse.html
Contact: hpgc@mendel.mgh.harvard.edu
```

CLONE LENGTH: This sequence represents the entire insert of this clone unless otherwise noted. If there are overlapping clones, the overlaps are noted in the beginning and end of the Features listing.

ANNOTATION OF FEATURES:

STSS are identified using ePCR (Genome Res. 7:541-550). Repeats are identified using RepeatMasker (A. Smit and P. Green, unpublished.) for Human and Mouse sequences.

Genes and Regions of sequence similarity are identified by BLAST (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to EST and cDNA sequences in Unigene. Genes demonstrate at least two exons flanked by consensus splice sites that maintain sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Attempts are made to complete double stranded sequence for all regions. All sequence is completed to a standard of coverage with a minimum of 3 reads with no ambiguities. If the sequence coverage for a region does not meet this standard, it is indicated in the annotation as Low Coverage. Low coverage linkages are verified by PCR product size verification or verification of forward and reverse reads from clones which span the low coverage area.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality standards - estimated average error rate is less than 1 per 10,000 bases using the Consed quality parameters. Regions that do not meet this requirement are annotated as Low Quality.

-----Summary Statistics

Center project name: AFC
Sequencing vector: pUC18; L08752
Chemistry: Dye-terminator Big Dye; 100%
Assembly program: Phrap version 0.990329
Contig length: 219559
Fraction of Phrap value < 40: 0.001207
Error Rate in Consed: 0.01 per 10,000 bases
Number of N's in consensus: 0

----- Distribution of Quality < 40 Bases:

1000
900
800

Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J., Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L., Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D., Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K., Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K., Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G., Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P., Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M., Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W., Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K., Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J., Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hognes,M., Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A., Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A., Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C., Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J., Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J., Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J., Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A., Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E., Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E., Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S., Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L., Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Nwaokemeleh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K., Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C., Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L., Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R., Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F., Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J., Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H., Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D., Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J., Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C., Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K., Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J., Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F., Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K., Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V., Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O., Weinstock,G. and Gibbs,R.A.

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

2 (bases 1 to 227533)
Worley,K.C.
Direct Submission

Submitted (15-SEP-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 227533)
Rat Genome Sequencing Consortium.
Direct Submission

REFERENCE
AUTHORS
TITLE
JOURNAL

Submitted (09-MAY-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
On May 9, 2003 this sequence version replaced gi:22856993.

COMMENT

The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GBQZ
Center clone name: CH230-6C1
----- Summary Statistics
Assembly program: Atlas;
Consensus quality: 220893 bases at least Q40
Consensus quality: 223841 bases at least Q30
Consensus quality: 225204 bases at least Q20
Estimated insert size: 232358; sum-of-contigs estimation
Quality coverage: 8x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently consists of 3 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.
* 1 224928: contig of 224928 bp in length
* 224929 225028: gap of unknown length
* 225029 226314: contig of 1286 bp in length
* 226315 226414: gap of unknown length
* 226415 227533: contig of 1119 bp in length.
FEATURES
source
1. .227533
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-6C1"
ORIGIN

Query Match 3.4%; Score 52.2; DB 2; Length 227533;
Best Local Similarity 47.4%; Pred. No. 0.027;
Matches 190; Conservative 0; Mismatches 208; Indels 3; Gaps 1;
QY 60 GCCTCTTCTGCTCCCATCTCGAGCATGCCAATGCCCTCTCTGCAACCCCATATAGCCAGCTCA 119
Db 177611 GCCTCAGCCTCAGCCCCAGCCTCAGCCCCAGCCTCAGCCTCAGGCCCGCAGCCTCAGCCTCA 177670
QY 120 GTACCCAGTCTCTCAGCCGCACACCCCTGCTCTCTCTCTGCAACCCCATATAGCCAGCTCGCC 179
Db 177671 GCCTCAGGCCAGCCTCAGCCCCAGCCTCAGCCTCAGCCCCAGCCTCAGCCCCAGCCCCA 177730
QY 180 CGCTCCTCACTCGTACATGGGGCAGCGCCGCTACCGGCTGATCTGAACAGGTACCCCGC 239
Db 177731 GCCTCAGCCTAGCCCCAGCCTCAGCCTCAGCCTCAGCCTCAGCCTCAGCCCCAGCCTCA 177790
QY 240 ATCAAGTCAGGATGTTTACGGCTTCTGCTGCGCCGATATGCCCCACACTACCGTGGG 299
Db 177791 GCCTCAGCCTCAGCCCCAGCCTCAGCCTCAGCCCCAGCCTCAGCCTCAGCCTCAGCCCCA 177850
QY 300 CAGCTTGCCTCCGACATCTTCT 359
Db 177851 GCCTCAGCCTCAGCCCCAGCCTCAGCCTCAGCCCCAGCCTCAGCCCCAGCCTCAGCCTCA 177910
QY 360 GCAATCGCCGCACTAT---CCT 416
Db 177911 GCCCCAGCCCCAGCCTCAGCCTCAGCCTCAACCCAGCCCCAGCCCCAGCCCCAGCCTCA 177970
QY 417 GTCGTACCCGAGCCAATTGCGCGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 457
Db 177971 GCCCCAGCCTCAGCGTCAGCCTCAGCCCCAGCCCCAGCCCCAGCCCC 178011
RESULT 14
AC122509
LOCUS
DEFINITION Mus musculus BAC clone RP24-443G20 from chromosome 16, complete sequence.
AC122509 178757 bp DNA linear ROD 27-NOV-2003
Mus musculus BAC clone RP24-443G20 from chromosome 16, complete sequence.

ACCESSION AC122509
VERSION AC122509.2 GI:30985117
KEYWORDS HTG.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
AUTHORS 1 (bases 1 to 178757)
TITLE Du, H., Boyer, E. and Creason, K.
JOURNAL The sequence of Mus musculus BAC clone RP24-443G20
REFERENCE 2 (bases 1 to 178757)
AUTHORS Wilson, R.
TITLE Sequencing of Mus musculus
JOURNAL Unpublished (2001)
REFERENCE 3 (bases 1 to 178757)
AUTHORS McPherson, J.D. and Waterston, R.H.
TITLE Direct Submission
JOURNAL Submitted (23-MAY-2002) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
REFERENCE 4 (bases 1 to 178757)
AUTHORS Wilson, R.K.
TITLE Direct Submission
JOURNAL Submitted (22-MAY-2003) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
REFERENCE 5 (bases 1 to 178757)
AUTHORS Wilson, R.
TITLE Direct Submission
JOURNAL Submitted (27-NOV-2003) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On May 22, 2003 this sequence version replaced gi:21105973.
----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: <http://genome.wustl.edu>
Contact: submissions@watson.wustl.edu
----- Summary Statistics
Center project name: M_BB0443G20

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. Wes Warren, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu>

SOURCE INFORMATION:

The RPCI-24 BAC library has been constructed by Pieter de Jong and coworkers (<http://www.chori.org>) from male C57BL/6J mouse spleen and/or brain genomic DNA. The clone and detailed information can be obtained from Pieter de Jong and coworkers at <http://www.chori.org>

NEIGHBORING SEQUENCE INFORMATION:

This sequence is the entire insert of the clone.

Location/Qualifiers
1. 178757
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="16"

FEATURES

source

repeat_region /map="16"
repeat_region /clone="RP24-443G20"
repeat_region /clone_lib="RPCI-24"
repeat_region 1123..1175
repeat_region /rpt_family="B4"
repeat_region 2269..2522
repeat_region /rpt_family="L1"
repeat_region 4511..4687
repeat_region /rpt_family="CR1"
repeat_region 5027..5985
repeat_region /rpt_family="L1"
repeat_region 6598..6703
repeat_region /rpt_family="MER103"
repeat_region 9601..9658
repeat_region /rpt_family="ERV1"
repeat_region 11560..11906
repeat_region /rpt_family="MaLR"
repeat_region 12066..12931
repeat_region /rpt_family="L1"
repeat_region 13001..13365
repeat_region /rpt_family="MaLR"
repeat_region 13640..14068
repeat_region /rpt_family="MaLR"
repeat_region 14376..14602
repeat_region /rpt_family="MIR"
repeat_region 17949..18060
repeat_region /rpt_family="B4"
repeat_region 18213..18359
repeat_region /rpt_family="Alu"
repeat_region 19039..19112
repeat_region /rpt_family="L2"
repeat_region 21794..21861
repeat_region /rpt_family="L1"
repeat_region 21930..21985
repeat_region /rpt_family="ERV1"
repeat_region 22019..22357
repeat_region /rpt_family="L1"
repeat_region 25268..25415
repeat_region /rpt_family="Alu"
repeat_region 25795..26039
repeat_region /rpt_family="B4"
repeat_region 26754..27200
repeat_region /rpt_family="ERV1"
repeat_region 27299..27440
repeat_region /rpt_family="B4"
repeat_region 30863..31039
repeat_region /rpt_family="MaLR"
repeat_region 31108..31319
repeat_region /rpt_family="MaLR"
repeat_region 31433..31532
repeat_region /rpt_family="Alu"
repeat_region 32245..32332
repeat_region /rpt_family="B4"
repeat_region 32333..32460
repeat_region /rpt_family="Alu"
repeat_region 32461..32522
repeat_region /rpt_family="B4"
repeat_region 32753..32982
repeat_region /rpt_family="L2"
repeat_region 34404..34484
repeat_region /rpt_family="B4"
repeat_region 34432..34628
repeat_region /rpt_family="Alu"
repeat_region 34676..34738
repeat_region /rpt_family="Alu"
repeat_region 34870..34987
repeat_region /rpt_family="B4"
repeat_region 37212..37345
repeat_region /rpt_family="L2"
repeat_region 40253..40414
repeat_region /rpt_family="L1"
repeat_region 40656..41180
repeat_region /rpt_family="L1"

```
repeat_region 41219. .41430
/rpt_family="MIR"
43486. .43794
/rpt_family="MaLR"
44215. .44393
/rpt_family="MaLR"
44768. .44929
/rpt_family="MaLR"
44965. .45091
/rpt_family="L1"
45491. .45725
/rpt_family="RMER3"
45532. .45725
/rpt_family="ERVK"
45737. .45919
/rpt_family="MaLR"
48222. .48429
/rpt_family="MER1_type"
49042. .49130
/rpt_family="MIR"
49499. .49761
/rpt_family="B4"
49988. .50332
/rpt_family="MaLR"
51110. .51439
/rpt_family="ERVK"
52948. .53013
/rpt_family="ID"
54787. .54987
/rpt_family="L1"
56208. .56318
/rpt_family="MIR"
57757. .57906
/rpt_family="B4"
58096. .58462
/rpt_family="L1"
59121. .59261
/rpt_family="MIR"
59814. .60375
/rpt_family="RMER19B"
61665. .61721
/rpt_family="ERV1"
63269. .63463
/rpt_family="MER1_type"
64122. .64281
/rpt_family="B2"
64288. .64337
/rpt_family="ERV1"
65623. .65853
/rpt_family="B4"
66706. .66860
```

```
Query Match 3.3%; Score 52; DB 10; Length 178757;
Best Local Similarity 44.6%; Pred. No. 0.031;
Matches 205; Conservative 0; Mismatches 255; Indels 0; Gaps 0;

QY 6 TCCTAGAAACCATCCCTCTCGGCTCCGCTCTACCGAGTCTGCGCTCAAGGATCGGCGGCTCT 65
Db 161448 TCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCC 161507

QY 66 TCCTTCTGCTCCCATCTCGAGCATGCCAATGCCTCAGTACACGATGCAGCCTCAGTACCC 125
Db 161508 TCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCC 161567

QY 126 AGTCTCTCAGCGGCACACCCCTGCTCCTCTGCAACCCCATCATAGCCAGTCCGCCGCTCC 185
Db 161568 TCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCC 161627

QY 186 TCACTCGTACATGGGGCAGCGCGCTACCGGCTGATCTGAACAGGTACCCCGCATCAAG 245
Db 161628 TCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCC 161687

QY 246 TCACGATGTTTACGGCTTCTGCTGCGCGGATGATGCCCACTACCGTGGGAGCTT 305
```

```
Db 161688 TCCTGCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCC 161747

QY 306 GCCTCCGACATCTTTCCTTTCTCATCCCAATCCGAGGCGCAGGCACAGGCGCAATC 365
Db 161748 TCCTCCTCCTCCTGCTCCTGCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTGCTGC 161807

QY 366 GCGCACTATCCTCCTCCTCATAGCGTGTCCGCGCCGCTTCCAGCGCTCAGTGTACCC 425
Db 161808 TCCTCCTGCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCC 161867

QY 426 GCAGCCAAATTGCGCGGCGCCTCCCGGGACCGTGTGCT 465
Db 161868 TCCTCCTGCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTGCT 161907
```

RESULT 15

AC128967

LOCUS

DEFINITION

AC128967

AC128967

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

AC128967 197714 bp DNA linear HTG 20-NOV-2002
Rattus norvegicus clone CH230-194J4, *** SEQUENCING IN PROGRESS
***, 3 unordered pieces.
AC128967 GI:25139175
HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
Rattus norvegicus (Norway rat)
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 197714)

Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
Biswal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Baves,K.,
Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hognes,M.,
Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
Nwaokemeleh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajls,D.,
Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 3, 2004, 05:01:29 ; Search time 611 Seconds
(without alignments)
10825.605 Million cell updates/sec

Title: US-10-029-180-7
Perfect score: 1557
Sequence: 1 atggatcctagaaacaccatcc.....ccgttcagcccgccagacaa 1557

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N Geneseq_29Jan04:*
1: Geneseqn1980s:*
2: Geneseqn1990s:*
3: Geneseqn2000s:*
4: Geneseqn2001as:*
5: Geneseqn2001bs:*
6: Geneseqn2002s:*
7: Geneseqn2003as:*
8: Geneseqn2003bs:*
9: Geneseqn2003cs:*
10: Geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	1557	100.0	1557	6	ABQ94220	Abq94220 FLO11 gen
2	1557	100.0	1800	6	ABQ94276	Abq94276 Fungal ge
3	1557	100.0	1833	6	ABQ94275	Abq94275 Fungal ge
4	391.6	25.2	867	7	ABZ51240	Abz51240 Aspergill
5	280.8	18.0	612	7	ABZ54133	Abz54133 Aspergill
6	59	3.8	236	3	AAC57129	Aac57129 Pinus rad
7	51	3.3	5452	9	ADC86736	Adc86736 Human GPC
8	49.2	3.2	1856	4	ABL28791	AbL28791 Drosophil
9	49.2	3.2	3219	4	ABL28790	AbL28790 Drosophil
10	49.2	3.2	6530	2	AAQ51557	Aaq51557 Lorocrin
11	49.2	3.2	6530	2	AAZ22072	Aaz22072 Nucleotid
12	49.2	3.2	6530	4	AAC68948	Aac68948 Mouse lor
13	48.6	3.1	3073	7	AAD51683	Aad51683 Human nuc
14	46.6	3.0	549	3	AAC93730	Aac93730 Cat flea
15	46.6	3.0	2000	7	ADA71938	Ada71938 Rice gene
16	46	3.0	110000	4	AAI99682_16	Continuation (17 o
17	46	3.0	110000	4	AAI99683_16	Continuation (17 o
18	45.2	2.9	492	2	AAT08579	Aat08579 Partial g
19	45.2	2.9	720	5	AAS05643	Aas05643 Mammalian
20	45.2	2.9	1476	2	AAT08578	Aat08578 Zinc fing
21	45	2.9	3133	9	ADC86738	Adc86738 Human GPC
22	44.6	2.9	3269	4	ABL26283	AbL26283 Drosophil
23	44.6	2.9	12920	4	ABL26282	AbL26282 Drosophil

24	44.4	2.9	4409	4	ABL17895	AbL17895 Drosophil
25	44.4	2.9	5021	4	ABL11783	AbL11783 Drosophil
26	44.4	2.9	7394	4	ABL17894	AbL17894 Drosophil
27	44.4	2.9	7614	4	ABL11782	AbL11782 Drosophil
c 28	44.4	2.9	10732	3	AAA10594	Aaa10594 Gene enco
c 29	44.4	2.9	31814	7	AAD47150	Aad47150 Human Ras
c 30	43.8	2.8	2480	4	ABL25810	AbL25810 Drosophil
31	43.8	2.8	2874	4	ABL28677	AbL28677 Drosophil
c 32	43.8	2.8	5125	4	ABL28676	AbL28676 Drosophil
33	43.2	2.8	646	4	AAS26107	Aas26107 Human cDN
34	43.2	2.8	646	7	ABX73448	Abx73448 Human nov
35	43	2.8	336	2	AAV86698	Aav86698 EST clone
36	43	2.8	1506	4	ABL29105	AbL29105 Drosophil
37	43	2.8	1696	7	ACC83475	Acc83475 Mouse cho
38	43	2.8	2336	3	AAZ35052	Aaz35052 Mouse CNR
39	43	2.8	2651	3	AAZ35053	Aaz35053 Mouse CNR
40	43	2.8	3215	4	ABL29108	AbL29108 Drosophil
41	43	2.8	3506	4	ABL29104	AbL29104 Drosophil
42	43	2.8	3847	3	AAZ35044	Aaz35044 Mouse CNR
43	43	2.8	4841	4	AAK52955	Aak52955 Human pol
44	43	2.8	4880	4	AAK51971	Aak51971 Human pol
45	43	2.8	4898	7	ABZ79896	Abz79896 Human nuc

ALIGNMENTS

RESULT 1
ABQ94220
ID ABQ94220 standard; DNA; 1557 BP.
XX
AC ABQ94220;
XX
DT 22-OCT-2002 (first entry)
XX
DE FLO11 gene expression regulator An09 coding sequence.
XX

KW Fungal gene expression regulator; fungicide; gene therapy; An01; An05;
KW An09; An13; An17; An20; An28; An34; At01-1; At01-2; At03; At05;
KW At07; At08; At11; At14; At16; At18; At19; At20; At22; At24; At27; At32;
KW PC05; PC06; PC07; PC08; PC09; PC10; PC18; PC23; PC24; PC25; PC33; PC34;
KW FLO11; fungal invasion; secondary metabolite; At279; At286; At291; At320;
KW At322; An1000; At167; At221; At233; At239; At240; At274; PC1000; PC1001;
KW lovF; lovE; lovastatin; PC804; acvA; penicillin; antifungal; gene; ds.
XX

OS Unidentified.

PN WO200257456-A2.

XX

PD 25-JUL-2002.

XX

PF 24-DEC-2001; 2001WO-US049911.

XX

PR 22-DEC-2000; 2000US-0257431P.

XX

PA (MICR-) MICROBIA.

XX

PI Cali BM, Madden KT, Milne TG, Zhang L, Silva JC, Trueheart J;

PI Holtzman D, Sherman A;

XX

DR WPI; 2002-627368/67.

DR P-PSDB; ABP63087.

XX

PT New isolated or recombinant gene, or purified protein, useful in
PT regulating fungal gene expression of FLO11, lovF, lovE or acvA for the
PT production of enzymes, secondary metabolites or other commercially and
PT medically useful products.

XX Claim 1; Page 36-37; 71pp; English.

PS

XX

CC The present invention relates to novel fungal gene expression regulators

CC (ABQ94217-ABQ94285 and ABP63084-ABP63152). An01, An05, An09, An10, An13,
CC An17, An20, An28, An34, At01-1, At01-2, At03, At05, At07, At08, At11,

CC At14, At16, At18, At19, At20, At22, At24, At27, At32, Pc05, Pc06, Pc07, Pc08, Pc09, Pc10, Pc18, Pc23, Pc24, Pc25, Pc33 and Pc34 are FLO11 gene expression regulators. FLO11 is required for fungal invasion and its expression is believed to be regulated by factors that also modulate secondary metabolite production. At279, At286, At291, At320, At322, An1000, At167, At221, At233, At239, At240, At274, Pc1000 and Pc1001 are lovF gene expression regulators, and At501 and At574 are lovE gene expression regulators. lovF and lovE are believed to be involved in the production of the secondary metabolite lovastatin. Pc804 is an acvA gene expression regulator. acvA is involved in the production of the secondary metabolite penicillin. The fungal gene expression regulators and their coding sequences are useful in regulating or manipulating the expression of fungal genes that are involved in the production of enzymes, secondary metabolites and other commercially and medically useful products, in order to achieve maximum benefit. The genes may also be used to identify genes relevant to fungal invasion which may act as targets for the development of antifungal drugs

XX

Sequence 1557 BP; 329 A; 533 C; 385 G; 310 T; 0 U; 0 Other;

Query Match 100.0%; Score 1557; DB 6; Length 1557;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1557; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGGATCCTAGAAACCATCCCTCTCGGCCCTCCGTACACAGTCTGCCTCAGGATCGGG 60
DB 1 ATGGATCCTAGAAACCATCCCTCTCGGCCCTCCGTACACAGTCTGCCTCAGGATCGGG 60

QY 61 CCTCTTCCTTCTGCTCCCATCTCGAGCATGCCAATGCCCTCAGTACACGATCAGCCTCAG 120
DB 61 CCTCTTCCTTCTGCTCCCATCTCGAGCATGCCAATGCCCTCAGTACACGATCAGCCTCAG 120

QY 121 TACCCAGTCTCTCAGCCGCACACCCCTGCCCTCCTCTGCAACCCCATCATAGCAGTCGCC 180
DB 121 TACCCAGTCTCTCAGCCGCACACCCCTGCCCTCCTCTGCAACCCCATCATAGCAGTCGCC 180

QY 181 GCTCCTCACTCGTACATGGGGCAGCCCGGTACCGGCTGATCTGAACAGGTACCCCGCA 240
DB 181 GCTCCTCACTCGTACATGGGGCAGCCCGGTACCGGCTGATCTGAACAGGTACCCCGCA 240

QY 241 TCAAGTCACGATGTTACGGCTCTTCTGCTGCGCGGATAATGCCCCACACATACCGTGGC 300
DB 241 TCAAGTCACGATGTTACGGCTCTTCTGCTGCGCGGATAATGCCCCACACATACCGTGGC 300

QY 301 AGCTTGCCTCCGACATCTTCTCTCATCCCAATCCGCGAGGCGCAGGCACAGGCGCAG 360
DB 301 AGCTTGCCTCCGACATCTTCTCTCATCCCAATCCGCGAGGCGCAGGCACAGGCGCAG 360

QY 361 CAATCGCCGCACTATCCTCCTCATAGCGTGCTCCCGGACCGCTTCCAGCGCTCAGTCG 420
DB 361 CAATCGCCGCACTATCCTCCTCATAGCGTGCTCCCGGACCGCTTCCAGCGCTCAGTCG 420

QY 421 TACCCGCGCAATTCGCGCGGCGCCTCCCGGACCGCTCGTGCTGCTGACTTCAACAATGGA 480
DB 421 TACCCGCGCAATTCGCGCGGCGCCTCCCGGACCGCTCGTGCTGCTGACTTCAACAATGGA 480

QY 481 CTTCCTTCAGGAGCATTCAGTTATTTCGGACGGAAGCCTCAAGTTGGACCCCGTTGCT 540
DB 481 CTTCCTTCAGGAGCATTCAGTTATTTCGGACGGAAGCCTCAAGTTGGACCCCGTTGCT 540

QY 541 GCGAATGGTGCTGCGCGGCTATCCCGGGAAGGACTCCCCCGGAACCCAGGTTGTTGTTCT 600
DB 541 GCGAATGGTGCTGCGCGGCTATCCCGGGAAGGACTCCCCCGGAACCCAGGTTGTTGTTCT 600

QY 601 CAGGGCGCAGCGGTATCCTTCGAGTGTTCGGGACGCGCAACTCCGGTCAACAATGGT 660
DB 601 CAGGGCGCAGCGGTATCCTTCGAGTGTTCGGGACGCGCAACTCCGGTCAACAATGGT 660

QY 661 GTTAACGGCACCGGCAAGACACTACTATCCCGGCCAAGGATGCCGATGGAAAGTCCCT 720
DB 661 GTTAACGGCACCGGCAAGACACTACTATCCCGGCCAAGGATGCCGATGGAAAGTCCCT 720

QY 721 TGCCCGAACTGTAAACAGACTTATCTTTCATGCCAAGCATCTCAAGCGCCATCTGCTACGC 780

DB 721 TGCCCGAACTGTAAACAGACTTATCTTTCATGCCAAGCATCTCAAGCGCCATCTGCTACGC 780

QY 781 CACACTGGTGACCGCCCGTACATGTGTCTTCTTGCAAAAGACACCTTCTCTCGCAGTGAT 840

DB 781 CACACTGGTGACCGCCCGTACATGTGTCTTCTTGCAAAAGACACCTTCTCTCGCAGTGAT 840

QY 841 ATCCTGAAACGTCATTTCCAAAAATGCTCAATCAGGCGTGGTAAACCCCAACCGGAGCAACG 900

DB 841 ATCCTGAAACGTCATTTCCAAAAATGCTCAATCAGGCGTGGTAAACCCCAACCGGAGCAACG 900

QY 901 CACTTGTCCGACCCCAATGCGCATGTGAAGAGGTCCCAACAGCAGGCTGCGGCGAATCCT 960

DB 901 CACTTGTCCGACCCCAATGCGCATGTGAAGAGGTCCCAACAGCAGGCTGCGGCGAATCCT 960

QY 961 GTAAACCTGTCCAGGATGAAGTCAAGTACGTCAGTACCGTCCCGCTCCCAATGGCATCCCGGC 1020

DB 961 GTAAACCTGTCCAGGATGAAGTCAAGTACGTCAGTACCGTCCCGCTCCCAATGGCATCCCGGC 1020

QY 1021 ACGACTTACGGCGAGGAGCCGTCAACGGCAATGGACTAGTCCGGCCCGCCAGGGTAC 1080

DB 1021 ACGACTTACGGCGAGGAGCCGTCAACGGCAATGGACTAGTCCGGCCCGCCAGGGTAC 1080

QY 1081 GCGGATCACAGACTATGGGCTTCCCAATGTCCATCGTCAACGGGATGGGCCGTGGTCAG 1140

DB 1081 GCGGATCACAGACTATGGGCTTCCCAATGTCCATCGTCAACGGGATGGGCCGTGGTCAG 1140

QY 1141 CCTGAAGACGCGTTTCCCGCGCGCCGCGCATCAAGGAGCCCTTGGCCACAAGCTCCC 1200

DB 1141 CCTGAAGACGCGTTTCCCGCGCGCCGCGCATCAAGGAGCCCTTGGCCACAAGCTCCC 1200

QY 1201 AAGCAGAGCCCGTATCTCGTGCAGCCGCGTGTGACCCCTTCTGGCCACCAGTTGAATATT 1260

DB 1201 AAGCAGAGCCCGTATCTCGTGCAGCCGCGTGTGACCCCTTCTGGCCACCAGTTGAATATT 1260

QY 1261 GACCGAAACATCGAGCAGGTAAACCAACCGGTTGTTCAAGACCCCAAGCCCTGTGATG 1320

DB 1261 GACCGAAACATCGAGCAGGTAAACCAACCGGTTGTTCAAGACCCCAAGCCCTGTGATG 1320

QY 1321 CCAGGACATCCCGGCCACCCCGGTGAGTTGACTGGACGTCATGTTCCAAACCTCAAGCT 1380

DB 1321 CCAGGACATCCCGGCCACCCCGGTGAGTTGACTGGACGTCATGTTCCAAACCTCAAGCT 1380

QY 1381 CCCGAGGCTACATGTTCTCCAGTCTATGCTGCTGGTCAAGAGCCCATCCACGCTCAT 1440

DB 1381 CCCGAGGCTACATGTTCTCCAGTCTATGCTGCTGGTCAAGAGCCCATCCACGCTCAT 1440

QY 1441 GTCGAGACCGGAAAGTATTACCCCAACACTACCGTGGTCAAGAGAGTGAATGAAC 1500

DB 1441 GTCGAGACCGGAAAGTATTACCCCAACACTACCGTGGTCAAGAGAGTGAATGAAC 1500

QY 1501 GGTCTCTATCTGGCTTCGACTATGAGTGGGACGGCACCGTTTCAGCCCGCCAGACAA 1557

DB 1501 GGTCTCTATCTGGCTTCGACTATGAGTGGGACGGCACCGTTTCAGCCCGCCAGACAA 1557

RESULT 2

ABQ94276

ID ABQ94276 standard; DNA; 1800 BP.

XX AC ABQ94276;

XX XX 22-OCT-2002 (first entry)

XX DE Fungal gene expression regulator coding sequence SEQ ID 118.

XX Fungal gene expression regulator; fungicide; gene therapy; An01; An05; An09; An10; An13; An17; An20; An28; An34; At01-1; At01-2; At03; At05; At07; At08; At11; At14; At16; At18; At19; At20; At22; At24; At27; At32; Pc05; Pc06; Pc07; Pc08; Pc09; Pc10; Pc18; Pc23; Pc24; Pc25; Pc33; Pc34; FLO11; fungal invasion; secondary metabolite; At279; At286; At291; At320; At322; An1000; At167; At221; At233; At239; At240; At274; Pc1000; Pc1001; lovF; lovE; lovastatin; Pc804; acvA; penicillin; antifungal; gene; ds.

XX Unidentified.
OS
XX
PN WO200257456-A2.
XX
PD 25-JUL-2002.
XX
PF 24-DEC-2001; 2001WO-US049911.
XX
PR 22-DEC-2000; 2000US-0257431P.
XX
PA (MICR-) MICROBIA.
XX
PI Cali BM, Madden KT, Milne TG, Zhang L, Silva JC, Trueheart J;
PI Holtzman D, Sherman A;
XX
DR WPI; 2002-627368/67.
XX
PT New isolated or recombinant gene, or purified protein, useful in
PT regulating fungal gene expression of FLO11, lovF, lovE or acvA for the
PT production of enzymes, secondary metabolites or other commercially and
PT medically useful products.
XX
PS Claim 67; Page 66-67; 71pp; English.
XX
CC The present invention relates to novel fungal gene expression regulators
CC (ABQ94217-ABQ94285 and ABP63084-ABP63152). An01, An05, An09, An10, An13,
CC An17, An20, An28, An34, At01-1, At01-2, At03, At05, At07, At08, At11,
CC At14, At16, At18, At19, At20, At22, At24, At27, At32, Pc05, Pc06, Pc07,
CC Pc08, Pc09, Pc10, Pc18, Pc23, Pc24, Pc25, Pc33 and Pc34 are FLO11 gene
CC expression regulators. FLO11 is required for fungal invasion and its
CC expression is believed to be regulated by factors that also modulate
CC secondary metabolite production. At279, At286, At291, At320, At322,
CC An1000, At167, At221, At233, At239, At240, At274, Pc1000 and Pc1001 are
CC lovF gene expression regulators, and At501 and At574 are love gene
CC expression regulators. lovF and lovE are believed to be involved in the
CC production of the secondary metabolite lovastatin. Pc804 is an acvA gene
CC expression regulator. acvA is involved in the production of the secondary
CC metabolite penicillin. The fungal gene expression regulators and their
CC coding sequences are useful in regulating or manipulating the expression
CC of fungal genes that are involved in the production of enzymes, secondary
CC metabolites and other commercially and medically useful products, in
CC order to achieve maximum benefit. The genes may also be used to identify
CC genes relevant to fungal invasion which may act as targets for the
CC development of antifungal drugs
XX
SQ Sequence 1800 BP; 372 A; 606 C; 463 G; 359 T; 0 U; 0 Other;

Query Match 100.0%; Score 1557; DB 6; Length 1800;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1557; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGGATCCTAGAAACCATCCCTCTCGGCCTCCGTCTACCAAGTCTGCCTCAAGGATCGGCG 60
Db |||||
244 ATGGATCCTAGAAACCATCCCTCTCGGCCTCCGTCTACCAAGTCTGCCTCAAGGATCGGCG 303

QY 61 CCTCTTCTCTGTCTCCCATCTCGAGCATGCCAATGCCTCAGTACACGATGCAGCCTCAG 120
Db |||||
304 CCTCTTCTCTGTCTCCCATCTCGAGCATGCCAATGCCTCAGTACACGATGCAGCCTCAG 363

QY 121 TACCCAGTCTCTCAGCCGCACACCCCTGCTCCTCTGCAACCCCATCATAGCCAGTCGCC 180
Db |||||
364 TACCCAGTCTCTCAGCCGCACACCCCTGCTCCTCTGCAACCCCATCATAGCCAGTCGCC 423

QY 181 GCTCCTCACTCGTACATGGGGCAGCCCGGTACCGGCCTGATCTGAACAGGTACCCCGCA 240
Db |||||
424 GCTCCTCACTCGTACATGGGGCAGCCCGGTACCGGCCTGATCTGAACAGGTACCCCGCA 483

QY 241 TCAAGTCAAGATGTTTACCGGTCTTCTGCTGCGCCGATATATGCCACACTACCGTGGGC 300
Db |||||
484 TCAAGTCAAGATGTTTACCGGTCTTCTGCTGCGCCGATATATGCCACACTACCGTGGGC 543

QY 301 AGCTTGCTCCGACATCTTTCTCTCATCCCAATCCGAGGCGGACAGGCACAGGCGCAG 360

Db ||||| AGCTTGCTCCGACATCTTTCTCTCATCCCAATCCGAGGCGGACAGGCGCAG 603
QY ||||| CAATCGCCGCACTATCTCTCTCATAGCGTGTCTCCGCGCGCTTCCAGCGCTCAGTCG 420
Db ||||| CAATCGCCGCACTATCTCTCTCATAGCGTGTCTCCGCGCGCTTCCAGCGCTCAGTCG 663
QY ||||| TACCCGAGCCAATTGCGCCGGCGCTCTCCGCGGACCGTCTGCTGCTCAACTTCAACAATGGA 480
Db ||||| TACCCGAGCCAATTGCGCCGGCGCTCTCCGCGGACCGTCTGCTGCTCAACTTCAACAATGGA 723
QY ||||| CTTCTTCAGGAGCATTCAGTTATTCGGACGGAAAGCCTCAAGSTTGGGACCCCGTTGCT 540
Db ||||| CTTCTTCAGGAGCATTCAGTTATTCGGACGGAAAGCCTCAAGSTTGGGACCCCGTTGCT 783
QY ||||| GCGAATGGTGCTGCGCCGTATCCCGGGAAGGACTCTCCCGGAAAGGTTGTTGTTCT 600
Db ||||| GCGAATGGTGCTGCGCCGTATCCCGGGAAGGACTCTCCCGGAAAGGTTGTTGTTCT 843
QY ||||| CAGGGCGACGCGGTATCCTTCGAGTGTTCGGGACCGGCAACTCCGGTCACAAATGGT 660
Db ||||| CAGGGCGACGCGGTATCCTTCGAGTGTTCGGGACCGGCAACTCCGGTCACAAATGGT 903
QY ||||| GTTAACGGCACCGGCAAGAACACTACTATCTCCGGCCAAGGATGCCGATGGAAAGTTCCCT 720
Db ||||| GTTAACGGCACCGGCAAGAACACTACTATCTCCGGCCAAGGATGCCGATGGAAAGTTCCCT 963
QY ||||| TGCCCGAACTGTAAACAAGACTTATCTTCAAGCATCTCAAGCGCCATCTGCTACGC 780
Db ||||| TGCCCGAACTGTAAACAAGACTTATCTTCAAGCATCTCAAGCGCCATCTGCTACGC 1023
QY ||||| CACACTGGTGACCGCCGTACATGTGTCTTTGCAAAAGACACCTTCTCTCGCAGTGAT 840
Db ||||| CACACTGGTGACCGCCGTACATGTGTCTTTGCAAAAGACACCTTCTCTCGCAGTGAT 1083
QY ||||| ATCTGAAACGTCATTTCCAAAATGCTCAATCAGGCGTGTAAACCCACCGGAGCAACG 900
Db ||||| ATCTGAAACGTCATTTCCAAAATGCTCAATCAGGCGTGTAAACCCACCGGAGCAACG 1143
QY ||||| CACTGTGCGACCCCAATGCGCATGTGAAGAGGTCCCAACAGCAGGCTGCGGCGAATCCT 960
Db ||||| CACTGTGCGACCCCAATGCGCATGTGAAGAGGTCCCAACAGCAGGCTGCGGCGAATCCT 1203
QY ||||| GTAAACCTGTCCAGGATGAAGTCAAGTAGTACCGTCCCGCTCCCAATGGCATCCCGGGC 1020
Db ||||| GTAAACCTGTCCAGGATGAAGTCAAGTAGTACCGTCCCGCTCCCAATGGCATCCCGGGC 1263
QY ||||| ACGACTTACGGCGAGGAGCCGCTCAACGGCAATGGACTAGTCCGGCCCGGCGGATC 1080
Db ||||| ACGACTTACGGCGAGGAGCCGCTCAACGGCAATGGACTAGTCCGGCCCGGCGGATC 1323
QY ||||| GCGGATCACGACACTATGGGCTTCCCAATGTCTCCGTCACCGGATGGCGCGTGTGTCAG 1140
Db ||||| GCGGATCACGACACTATGGGCTTCCCAATGTCTCCGTCACCGGATGGCGCGTGTGTCAG 1383
QY ||||| CCTGAAGACGCGTTTCCCGGCGGCGCGCGCATCAAGGAGCCCTTGGCCACAAGCTCCC 1200
Db ||||| CCTGAAGACGCGTTTCCCGGCGGCGCGCGCATCAAGGAGCCCTTGGCCACAAGCTCCC 1443
QY ||||| AAGCAGAGCCCGTATCTCTGTCAGCCGGGTGCTGACCCCTTCTGGCCACCAGTTGAATAT 1260
Db ||||| AAGCAGAGCCCGTATCTCTGTCAGCCGGGTGCTGACCCCTTCTGGCCACCAGTTGAATAT 1503
QY ||||| GACCGAAACATCGAGCAGGTAAACCAACCGGTTGTTCAAGACCCCAAGCCCTGTGATG 1320
Db ||||| GACCGAAACATCGAGCAGGTAAACCAACCGGTTGTTCAAGACCCCAAGCCCTGTGATG 1563
QY ||||| CCAGGACATCCCGGCCACCCCGGTGAGCTTGACTGGACGCTATGTTCCAACTCAAGCT 1380
Db ||||| CCAGGACATCCCGGCCACCCCGGTGAGCTTGACTGGACGCTATGTTCCAACTCAAGCT 1623
QY ||||| CCGAGGGCTACATGTTCTCCAGTCTATGCCTGGTGGTCAAGAGCCCATCCACGCTCAT 1440

Db 1624 CCCGAGGCTACATGTTCTCCAGTCTATGCCCTGGTGGTCAAGAGCCCATCCACGCTCAT 1683

QY 1441 GTCGAGACCGAGCGAAAGTATTACCCACCACTACCGCTGGTCAAGAGAGTGGAATGAAC 1500

Db 1684 GTCGAGACCGAGCGAAAGTATTACCCACCACTACCGCTGGTCAAGAGAGTGGAATGAAC 1743

QY 1501 GGCTCTATCTGGCTTCGACTATGAGTGGCGACGGCACCGTTTCAGCCCGCCAGACAA 1557

Db 1744 GGCTCTATCTGGCTTCGACTATGAGTGGCGACGGCACCGTTTCAGCCCGCCAGACAA 1800

RESULT 3

ID ABQ94275

XX ABQ94275 standard; DNA; 1833 BP.

AC ABQ94275;

XX

XX

DT 22-OCT-2002 (first entry)

XX

DE Fungal gene expression regulator coding sequence SEQ ID 117.

XX

KW Fungal gene expression regulator; fungicide; gene therapy; An01; An05;

KW An09; An10; An17; An20; An28; An34; At01-1; At01-2; At03; At05;

KW At07; At08; At11; At14; At16; At18; At19; At20; At22; At24; At27; At32;

KW Pc05; Pc06; Pc07; Pc08; Pc09; Pc10; Pc18; Pc23; Pc24; Pc25; Pc33; Pc34;

KW FLO11; fungal invasion; secondary metabolite; At279; At286; At291; At320;

KW At322; An1000; At167; At221; At233; At239; At240; At274; Pc1000; Pc1001;

KW lovF; lovE; lovastatin; Pc804; acvA; penicillin; antifungal; gene; ds.

XX

OS Unidentified.

XX

PN WO200257456-A2.

XX

XX 25-JUL-2002.

PD

XX

PF 24-DEC-2001; 2001WO-US049911.

XX

XX 22-DEC-2000; 2000US-0257431P.

PR

XX

PA (MICR-) MICROBIA.

XX

PI Cali BM, Madden KT, Milne TG, Zhang L, Silva JC, Trueheart J;

PI Holtzman D, Sherman A;

XX

DR WPI; 2002-627368/67.

XX

PT New isolated or recombinant gene, or purified protein, useful in

PT regulating fungal gene expression of FLO11, lovF, lovE or acvA for the

PT production of enzymes, secondary metabolites or other commercially and

PT medically useful products.

XX

PS Claim 67; Page 66; 71pp; English.

XX

CC The present invention relates to novel fungal gene expression regulators

CC (ABQ94217-ABQ94285 and ABP63084-ABP63152). An01, An05, An09, An10, An13,

CC An17, An20, An28, An34, At01-1, At01-2, At03, At05, At07, At08, At11,

CC At14, At16, At18, At19, At20, At22, At24, At27, At32, Pc05, Pc06, Pc07,

CC Pc08, Pc09, Pc10, Pc18, Pc23, Pc24, Pc25, Pc33 and Pc34 are FLO11 gene

CC expression regulators. FLO11 is required for fungal invasion and its

CC expression is believed to be regulated by factors that also modulate

CC secondary metabolite production. At279, At286, At291, At320, At322,

CC An1000, At167, At221, At233, At239, At240, At274, Pc1000 and Pc1001 are

CC lovF gene expression regulators, and At501 and At574 are lovE gene

CC expression regulators. lovF and lovE are believed to be involved in the

CC production of the secondary metabolite lovastatin. Pc804 is an acvA gene

CC expression regulator. acvA is involved in the production of the secondary

CC metabolite penicillin. The fungal gene expression regulators and their

CC coding sequences are useful in regulating or manipulating the expression

CC of fungal genes that are involved in the production of enzymes, secondary

CC metabolites and other commercially and medically useful products, in

CC order to achieve maximum benefit. The genes may also be used to identify

CC genes relevant to fungal invasion which may act as targets for the

CC development of antifungal drugs

Db 1237 GTAAACCTGTCCAGGATGAAGTCAGTAGTACCGTCCCGCTCCCAATGGCATCCCGGGC 1296

QY 1021 ACGACTTACGGGAGGAGCCGTCACGGCAATGGACTAGCTCCGGCCCCGAGGGTAC 1080

Db 1297 ACGACTTACGGGAGGAGCCGTCACGGCAATGGACTAGCTCCGGCCCCGAGGGTAC 1356

QY 1081 GCGGATCACAGACTATGGGCTTCCCAATGTCATCCGTCACAGGGATGGCCGTGTCAG 1140

Db 1357 GCGGATCACAGACTATGGGCTTCCCAATGTCATCCGTCACAGGGATGGCCGTGTCAG 1416

QY 1141 CCTGAAGACCGGTTTCCGGCGGCGCGGATCAAGGAGCCCTTGGCCACAAGCTCCC 1200

Db 1417 CCTGAAGACCGGTTTCCGGCGGCGGCGGATCAAGGAGCCCTTGGCCACAAGCTCCC 1476

QY 1201 AAGCAGAGCCCGTATCTCGTGCAGCCGGGTGCTGACCCCTTCTGGCCACCAGTTGAATATT 1260

Db 1477 AAGCAGAGCCCGTATCTCGTGCAGCCGGGTGCTGACCCCTTCTGGCCACCAGTTGAATATT 1536

QY 1261 GACCGAAACATCGAGCAGGTAAAAACAACCGGTTGTTCAAGACCCCAAGCGCCCTGTGATG 1320

Db 1537 GACCGAAACATCGAGCAGGTAAAAACAACCGGTTGTTCAAGACCCCAAGCGCCCTGTGATG 1596

QY 1321 CCAGGACATCCCGGCCACCCCGGTGAGCTTGACTGGACGTCTATGTTCCAAACCTCAAGCT 1380

Db 1597 CCAGGACATCCCGGCCACCCCGGTGAGCTTGACTGGACGTCTATGTTCCAAACCTCAAGCT 1656

QY 1381 CCCGAGGGCTACATGTTCTCCAGTCTATGCCTGGTCAAGAGCCCATCCACGCTCAT 1440

Db 1657 CCCGAGGGCTACATGTTCTCCAGTCTATGCCTGGTCAAGAGCCCATCCACGCTCAT 1716

QY 1441 GTCGAGACCGAGGAAAGTATTACCCCACTACCGCTGGTCAAGAGAGTGAATGAAC 1500

Db 1717 GTCGAGACCGAGGAAAGTATTACCCCACTACCGCTGGTCAAGAGAGTGAATGAAC 1776

QY 1501 GGTCCTCTATCTGGCTTCGACTATGAGTGGCGACGGCACCGTTTCAGCCCGCCAGACAA 1557

Db 1777 GGTCCTCTATCTGGCTTCGACTATGAGTGGCGACGGCACCGTTTCAGCCCGCCAGACAA 1833

RESULT 4

ABZ51240

ID ABZ51240 standard; cDNA; 867 BP.

AC ABZ51240;

XX 28-MAR-2003 (first entry)

XX Aspergillus oryzae polynucleotide SEQ ID NO 353.

DE Aspergillus oryzae; fermentation; fungus; industrial; EST;

XX expressed sequence tag; gene; ss.

KW Aspergillus oryzae.

XX WO200279476-A1.

PN 10-OCT-2002.

XX 22-MAR-2002; 2002WO-IB000890.

PF 30-MAR-2001; 2001JP-00098371.

XX (NAAD-) NAT INST ADVANCED IND SCI & TECHNOLOGY.

PA (NARE-) NAT RES INST BREWING.

PA (NORQ) NAT FOOD RES INST MIN AGRIC.

XX Machida M, Akita O, Kashiwagi Y, Kitamoto K, Horiuchi H;

PI Takeuchi M, Kobayashi T, Kitamoto N, Gomi K, Abe K;

XX WPI; 2003-046817/04.

XX Detection of expression of specific Aspergillus genes for monitoring the

PT fermentation and growth conditions of the fungus, using DNA probes.

XX Claim 1; SEQ ID NO 353; 48pp + Sequence Listing; Japanese.

PS The invention relates to a polynucleotide having any of 6006 specific

XX sequences (ABZ50888-ABZ56893), which are expressed by a fungus under

CC specific culture conditions including one or more of eutrophic,

CC oligotrophic, solid, early germination, alkaline, high temperature, low

CC temperature or maltose culture or polynucleotides stringently hybridising

CC to these sequences. The polynucleotides are useful for monitoring the

CC progress of fermentation and the growth conditions of a fungus,

CC especially of Aspergillus oryzae which is widely used in industrial

CC fermentation. Also monitoring for fungal contamination. Note: The

CC sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from WIPO

CC at ftp.wipo.int/pub/published_pct_sequences

XX Sequence 867 BP; 188 A; 257 C; 234 G; 182 T; 0 U; 6 Other;

SQ Query Match 25.2%; Score 391.6; DB 7; Length 867;

Best Local Similarity 71.6%; Pred. No. 1e-103;

Matches 591; Conservative 1; Mismatches 206; Indels 27; Gaps 5;

QY 751 GCCAAGCATCTCAAGCGCCATCTGCTACGCCACACACTGGTGACCCCGGTACATGTGTGT 810

Db 31 GGCAACATCTGAAGCGCCATCTGCTCASACATACCGGTGACCGCCCATACATGTGTGTN 90

QY 811 CTTTGAAAGACACCTTCTCTCGCAGTGATATCTTGAAACGTATTTCCAAAAATGCTCA 870

Db 91 CTGTGCAAGGATACTTTCTCCCGTAGTGATATTCTGAAGCGCCACTTCCAGAAGTGTTCG 150

QY 871 ATCAGGGGTGGTAACCCACCGGAGCAACGCACATTGTGCAACCCCAATGCGCATGTGAAG 930

Db 151 TTACGACGGGGCAACCCGACGGGAGCGACCCACTTGTGCAATCCCGAGGCACATTTTGAAG 210

QY 931 AGGTCCCAACAGCAGGCTGCGGCGAATCCTGTAAACCTGTCCAGGATGAAGTCAGTAGT 990

Db 211 AGGTCTCAA-----GCTGCGAATCCTGCTAAACCGGTTCAGGATGAAGTCAGTAGT 261

QY 991 ACCGTCCCGCTCCCAATGGCATCCCGGCACGACTTACGGCGAGGGAGCCGTCACCGGC 1050

Db 262 ACCGTCTCTCCGCCACTGCCCTTCCGGGTACGGCTTACGGCGAGGGGCCGTGAACGGT 321

QY 1051 AATGGACTAGCTCCGGCCCGGCAGGGTACGGGATCACCGAGTATGGCTTCCCAATG 1110

Db 322 AACGGGTGGCTTCAGGCCGACCTGGGTTACGGATCAGCAGCTCTGGGCTATCCGATG 381

QY 1111 TCATCCGTCAACGGGATGGCCCGTGTGTCAGCCTGAAGACGCGTTTCCCGGCGCGGCCG 1170

Db 382 TCGTCGGTCAATGGGATGAACCGTGGTCAACCGGACGATGCGTTTCCCGCTGGTCAGCCA 441

QY 1171 CATCAAGGAGCCCTTGGCCACAAGCTCCCAAGCAGAGCCCCGTATCTCGTCAGCCGGGT 1230

Db 442 CATCAGAGAGGCCCTTGGCTGGCGCTCCCAAGCAGAACCCGTATCTGGCGCAACCTGGC 501

QY 1231 GCTGACCCCTTCTGGCCACCAGTTGAAATATTGACCGGAACATCGAGCAGGTAAAAACAACCG 1290

Db 502 ACTGATGTATCTCAGCAGCTGAGTGTGACCGTCTTCCCTATGAACAGGTAAAGCCCCCG 561

QY 1291 GTTGTTCAAGACCCCAAGCGCCCTGTGATGCCAGGAC--ATCCCGGCCACCCCGGTGAG 1347

Db 562 GTGTTGCAAGACCCTAAGCGCCCAAGTAATGCCTGGGCCCCACTCCCAATCACACTGGCGAA 621

QY 1348 CTTGACTGGACGTCTATGTTCCAACTCAAGCTCCCGAGGGCTACA-----TGTTTC 1398

Db 622 ATTGACTGGACTTCGATGTTTTCAGCCCGGAGCGGAATGATGGTTACATCAACCCCGTTTTC 681

QY 1399 TCCCAGTCTATGCTGGTGGTCAAGAGCCCATCCACGCTCATGTGAGACCGAGCGAAAG 1458

Db 682 CCCCATCCATGGCGTCTGCGCCAGGAGCCGATCCATGCTCAGTCGATACCGAACCGGAAA 741

QY 1459 TATTACCCCACTACCGCTGGT---CAAGAGAGTGAATGAACGGTCTCTATCTGGCT 1515

Db 742 TTCTATCCTACCACAACCGGTGGTCCCCAGGAAGGTGGCATGATGATGTTTACTTGGCT 801

QY 1516 TCG---ACTATGAGTGGCGACGGCACCGGTCAGCCCGCCAGACAA 1557
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 802 TCGACCACTTGGCCGGTGATGGTACCGTCCAGCCCGGAGGCAA 846
| | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 5
ABZ54133
ID ABZ54133 standard; cDNA; 612 BP.
XX
AC ABZ54133;
XX
DT 28-MAR-2003 (first entry)
XX
XX Aspergillus oryzae polynucleotide SEQ ID NO 3246.
DE
XX
XX Aspergillus oryzae; fermentation; fungus; industrial; EST;
KW expressed sequence tag; gene; ss.
XX
OS Aspergillus oryzae.
XX
PN WO200279476-A1.
XX
XX
PD 10-OCT-2002.
XX
PF 22-MAR-2002; 2002WO-IB000890.
XX
PR 30-MAR-2001; 2001JP-00098371.
XX
XX (NAAD-) NAT INST ADVANCED IND SCI & TECHNOLOGY.
PA (NARE-) NAT RES INST BREWING.
PA (NORQ) NAT FOOD RES INST MIN AGRIC.
XX

PI Machida M, Akita O, Kashiwagi Y, Kitamoto K, Horiuchi H;
PI Takeuchi M, Kobayashi T, Kitamoto N, Gomi K, Abe K;
XX
DR WPI; 2003-046817/04.
XX
XX Detection of expression of specific Aspergillus genes for monitoring the
PT fermentation and growth conditions of the fungus, using DNA probes.
XX
PS Claim 1; SEQ ID NO 3246; 48pp + Sequence Listing; Japanese.
XX

CC The invention relates to a polynucleotide having any of 6006 specific
CC sequences (ABZ50888-ABZ56893), which are expressed by a fungus under
CC specific culture conditions including one or more of eutrophic,
CC oligotrophic, solid, early germination, alkaline, high temperature, low
CC temperature or maltose culture or polynucleotides stringently hybridising
CC to these sequences. The polynucleotides are useful for monitoring the
CC progress of fermentation and the growth conditions of a fungus,
CC especially of Aspergillus oryzae which is widely used in industrial
CC fermentation. Also monitoring for fungal contamination. Note: The
CC sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 612 BP; 108 A; 245 C; 121 G; 138 T; 0 U; 0 Other;

Query Match 18.0%; Score 280.8; DB 7; Length 612;
Best Local Similarity 75.6%; Pred. No. 2.1e-71;
Matches 393; Conservative 0; Mismatches 112; Indels 15; Gaps 3;

QY 1 ATGGATCCTAGAAACCATCCCTCTCGGCTCCGCTACACAGTCTGCTCAAGGATCGGCG 60
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 108 ATGGATCCCAGATCTCATCCGTCACGGCCCCCATCCACGCTTACCCCAAGGCTCCACG 167
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 61 CCTCTTCCTTCTGCTCCCATCTCGAGCATGCCAATGCCTCAGTACAGCATGCAGCTCAG 120
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 168 CCTCTCCCGACTACT---ATCTCGAGTATGCCAATGCCTCAGTACACCATGCAGGCTCAG 224
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 121 TACCCAGTCTTCAGCCGCACACCCCTGCTCCTCTGTGCAACCCCATCATAGCCAGTCGCC 180
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 225 TACCCTGTGCCAGCCTCATCTCTCCCGCGTGCAGCCTCATACAGCCCAATCTCCT 284
| | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 181 GCTCCTCACTCGTACATGGGCGAGCCCGCGTACCGGCTGATCTGAACAGGTACCCCGCA 240
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 285 GCCCCTCATTCCTACATGAGCAGCCCGCGGTACAGGCCGGATCTGTCGAGGTATCCCACT 344
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 241 TCAAGTCACGATGTTTACGCGTCTTCTGCTGCGCCGATATGCCCCACACTACCGTGGGC 300
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 345 TCTACTCAGATGTATACGCGTCTGCTACCGTCTCTATCATGCTTACACACCGTGGGT 404
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 301 AGCTTGCTCCGACATCTTCTCTTCTCATCCCAATCCGAGGCGCAGGCAGCGCAG 360
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 405 AGCCTGCTCCATCTTCTGTTCTCTCTCCACCCCAATCCGAGGCGAGCCCGAG----- 458
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 361 CAATCGCGCGCACTATCCTCCTCTCATAGGCTGCTCCCGCCGCTTCCAGCGCTCAGTCG 420
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 459 ---CCTTCGCAATATCTCCACCTCACAGCGTGTGCTCGGCTTCTAGCGCGCAGACA 515
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 421 TACCCGCGAGCCAATTGCGCGGCGCCTCCCGGGACCGTGTGCTGACTTCAACAATGA 480
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 516 TATCCGAGCCGATCGCGCGGACACCCCGGACCGCGCGGACTT---TTCTGGC 572
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 481 CTTCTTCAGGAGCATTCAGTTATTTCGGACGGAAAGCCTC 520
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 573 ATGCTTCCGGTGCTTTTCAGCTACTCAGATGGCAGGGGCTC 612
| | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 6
AAC57129
ID AAC57129 standard; DNA; 236 BP.
XX
AC AAC57129;
XX

DT 25-JAN-2001 (first entry)

XX Pinus radiata transcription factor DNA sequence #575.

DE
XX
KW plant; transcription factor; gene expression; eucalyptus; pine; acacia;
KW poplar; sweetgum; teak; mahogany; bZIP; G-box binding factor;
KW basic helix-loop-helix zipper; homeotic; homeodomain; homeobox; MADS;
KW homeodomain zipper; LIM domain; AP2; EREBs; zinc finger domain;
KW type 2 Cys2His2; CCAAT box element; MYB; ss.
XX

OS Pinus radiata.

XX WO200053724-A2.

PD 14-SEP-2000.

XX 09-MAR-2000; 2000WO-US006112.

XX 11-MAR-1999; 99US-00266513.

PR 18-AUG-1999; 99US-0149485P.

XX (GENE-) GENESIS RES & DEV CORP LTD.

PA (FLET-) FLETCHER CHALLENGE FORESTS LTD.

XX Wood M, Mcgrath A, Shenk MA, Glenn M;

XX WPI; 2000-579369/54.

PT New isolated polynucleotide encoding a plant transcription factor for
PT producing a plant e.g. a woody plant, preferably eucalyptus or pine,
PT having modified gene expression or modified activity of a polypeptide.

XX Claim 1; Page 592; 747pp; English.

CC The present invention relates to novel plant transcription factors from
CC Eucalyptus grandis or Pinus radiata. The present sequence is the coding
CC sequence for one such transcription factor. The transcription factor may
CC be used to produce a plant having modified gene expression such as a
CC woody plant e.g. a eucalyptus, pine, acacia, poplar, sweetgum, teak, or
CC mahogany species or to modify the activity of a polypeptide in a plant.
CC The transcription factors of the present invention are members from the

CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
CC ABB72072). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences

XX
SQ Sequence 1856 BP; 532 A; 451 C; 449 G; 424 T; 0 U; 0 Other;
Query Match 3.2%; Score 49.2; DB 4; Length 1856;
Best Local Similarity 59.2%; Pred. No. 0.0016;
Matches 84; Conservative 0; Mismatches 58; Indels 0; Gaps 0;

QY 714 GTTCCCTTGCCGAACTGTAACAAGACTTATCTTCATGCCAAGCATCTCAAGCGCCATCT 773
Db GTACTCCTGCAAAATGTGCAGCAAGACCTTTGTGCAGTCCACGACCTTAAGATCCATAT 1266

QY 774 GCTACGCCACACTGGTGACCGCCCGTACATGTGTCTTTGAAAGACACCTTCTCTCG 833
Db GCGGCGCACACGGCGGAGCGACCATATCAGTGTGGAGTGTGCGCGAGAGCTTTGTGTG 1326

QY 834 CAGTGATATCCTGAAACGTCAT 855
Db CGGTTCCGACCTGAATATCCAT 1348

RESULT 9
ABL28790/c
ID ABL28790 standard; DNA; 3219 BP.
XX
AC ABL28790;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 37843.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ds.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US009231.
XX
PR 23-MAR-2000; 2000US-0191637P.
PR 11-JUL-2000; 2000US-00614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signaling and cell-cell
PT interactions.
XX
PS Claim 1; SEQ ID NO 37843; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
CC ABB72072). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences

CC ABB72072). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences

XX
SQ Sequence 3219 BP; 842 A; 747 C; 731 G; 899 T; 0 U; 0 Other;
Query Match 3.2%; Score 49.2; DB 4; Length 3219;
Best Local Similarity 59.2%; Pred. No. 0.0021;
Matches 84; Conservative 0; Mismatches 58; Indels 0; Gaps 0;

QY 714 GTTCCCTTGCCGAACTGTAACAAGACTTATCTTCATGCCAAGCATCTCAAGCGCCATCT 773
Db GTACTCCTGCAAAATGTGCAGCAAGAGCTTTGTGCAGTCCACGACCTTAAGATCCATAT 901

QY 774 GCTACGCCACACTGGTGACCGCCCGTACATGTGTCTTTGAAAGACACCTTCTCTCG 833
Db GCGGCGCACACGGCGGAGCGACCATATCAGTGTGGAGTGTGCGCGAGAGCTTTGTGTG 841

QY 834 CAGTGATATCCTGAAACGTCAT 855
Db CGGTTCCGACCTGAATATCCAT 819

RESULT 10
AAQ51557/c
ID AAQ51557 standard; DNA; 6530 BP.
XX
AC AAQ51557;
XX
DT 25-MAR-2003 (revised)
DT 17-MAY-1994 (first entry)
XX
DE Loricrin gene.
XX
KW Loricrin; gene; constitutive; inducible; vector; 5' flanking region;
KW intron; 3' flanking region; TATA box; cap site; intron/exon boundary;
KW DNA cassette; linker; restriction site; induction; human; psoriasis;
KW epidermal cell; wound healing; ulcers; growth factor; antisense RNA;
KW transforming growth factor beta; cytokine; skin; tracheal epithelia;
KW cancer; E6; E7; papilloma virus; p53 protein; vaccine; oral epithelia;
KW capsid protein; squamous epithelium; viral; oesophageal epithelia;
KW vaginal epithelia; corneal epithelia; transduction; transplant; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT misc_RNA 1..1540
FT /*tag= a
FT /note= "5' flanking region"
FT intron 1587..2677
FT /*tag= b
FT /note= "Intron and intron/exon boundary"
FT misc_RNA 4384..6530
FT /*tag= c
FT /note= "3' flanking region"
XX
PN WO9322431-A1.
XX
PD 11-NOV-1993.
XX
PF 28-APR-1993; 93WO-US003993.
XX
PR 30-APR-1992; 92US-00876286.
XX
PA (BAYU) BAYLOR COLLEGE MEDICINE.
XX
PI Roop DR, Rothnagel JA, Greenhalgh DA;
XX
DR WPI; 1993-368790/46.
XX
PT Constitutive and inducible vectors based on loricrin or keratin K6 genes
PT - for selective expression in the epidermis, e.g. for wound healing,
PT treating psoriasis and skin cancer, etc.

XX Claim 23; Page 26-32; 68pp; English.

XX This sequence represents the loricrin gene and was used in the

CC construction of the loricrin constitutive vector of the invention. The

CC loricrin vector has a 5' flanking region of approx. 1.5 kb, an intron of

CC approx. 1.1 kb and a 3' flanking region of approx. 1.5 kb. The 5'

CC flanking region of the loricrin gene comprises a TATA box, cap site,

CC first intron and intron/exon boundary, all in the correct order and

CC position for expression of a nucleic acid cassette. The loricrin vector

CC also comprises a 3' flanking region of the loricrin gene and a linker

CC with a unique restriction site at the position of the start and stop

CC codons, joining the two flanking sequences and also having a position for

CC insertion of the DNA cassette (see also AAQ51558). Vectors such as this

CC can be used for; (a) in vitro induction of wounds, surgical incisions and

CC especially for stimulating healing of wounds, dermal cells, and

CC ulcers, (where the vector expresses a growth factor), (b) treating

CC psoriasis, (the vector expressing transforming growth factor beta, a

CC soluble cytokine or antisense RNA), (c) treating skin cancer, (the vector

CC expressing antisense RNA of the E6 or E7 gene of human papilloma virus or

CC normal p53 protein), (d) for vaccination (the vector expresses a viral

CC capsid protein, especially of human papilloma virus), or (e) treating

CC cancer of the squamous epithelium (of epidermal, oral, esophageal,

CC vaginal, tracheal or corneal epithelia, the vector expresses antisense

CC RNA). Alternatively human epithelial cells are transduced ex vivo, then

CC transplanted. (Updated on 25-MAR-2003 to correct PN field.)

XX

SQ Sequence 6530 BP; 1635 A; 1507 C; 1696 G; 1692 T; 0 U; 0 Other;

Query Match 3.2%; Score 49.2; DB 2; Length 6530;

Best Local Similarity 46.1%; Pred. No. 0.0029;

Matches 165; Conservative 0; Mismatches 193; Indels 0; Gaps 0;

QY 96 GCCTCAGTACAGGATGCGCTCAGTACCGAGTCTCTCAGCGGACACCGTGCCTCTCT 155

Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

3530 GCCTCCGTAGTCTGGCACTGATACTGTTGGAGCAGTCCGCGCGCCACCGGAGGA 3471

QY 156 GCAACCCCATCATAGCCAGTCCGCGCTCCTCACTCGTACATGGGGCAGCGCGTACCG 215

Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

3470 ACCACCTCCGAGCTCGTGCCGCGCCACCGGAATAGCCGCGCGCGCAGCTGGA 3411

QY 216 GCCTGATCTGAACAGGTACCCCGCATCAAGTCACGATGTTTACGGTCTTCTGTCGCC 275

Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

3410 TCCACCGCCAGCTGAGTAGCCGCGCCACAGCTAGAGCCGCGCTCCAGAGAGCCGCC 3351

QY 276 GATAATGCCCCACACTACCGTGGGAGCTGCTCCGACATCTTTCTTCTCATCCAA 335

Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

3350 GGAATAGCCGCTCCGAGCTGGAGCCACCGCGCCACAGAGTACTTGCCGCTCCGGA 3291

QY 336 TCCGAGGCGGAGGACAGGCAATCGCGCACTATCCTCCTCATAGCGTCT 395

Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

3290 ACCGCGCGCAGCCAGCTGCGGCTCCAGAGAGCCTCCGAGTAGGAGCCGCC 3231

QY 396 CCGCGCGCTTCCAGCGCTCAGTCGTACCGGAGCCAAATTGCGCGCGCTCCCGG 453

Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

3230 GCCTCCTGATCCGCTCCGAGCTGGAGCGCGCGCGCGCGCGCGCGCGCGCGCG 3173

RESULT 11

AAZ22072/c

ID AAZ22072 standard; DNA; 6530 BP.

XX AAZ22072;

AC AAZ22072;

XX 25-NOV-1999 (first entry)

DT

XX Nucleotide sequence of the loricrin gene.

DE

XX Vector; epidermal cell; loricrin gene; K6 keratin gene; wound;

KW surgical incision; disease; psoriasis; cancer; alopecia;

KW transgenic animal; ss.

XX

OS Mus sp.

XX US5958764-A.

PN

XX 28-SEP-1999.

PD

XX 01-NOV-1993; 93US-00146930.

PF

XX 30-APR-1992; 92US-00876286.

PR

XX 29-OCT-1993; 93US-00145388.

PA (BAYU) BAYLOR COLLEGE MEDICINE.

XX

XX Greenhalgh DA, Rothnagel JA, Roop DR;

PI WPI; 1999-561044/47.

XX

XX Gene therapy vectors for expression in the epidermis.

PT

XX Example 1; Col 25-32; 34pp; English.

PS

XX The specification describes a vector for the expression of a nucleic acid

CC sequence in an epidermal cell, where the nucleic acid sequence is not a

CC mammalian loricrin gene or a mammalian K6 keratin gene. The vector

CC comprises a 5' flanking region including the sequences required for its

CC expression; a 3' region from a mammalian loricrin or a mammalian K6

CC keratin gene which regulates expression (predominantly in an epidermal

CC cell) of the nucleic acid sequence with which it is associated; and a

CC linker connecting the 5' flanking region to the nucleic acid sequence,

CC where the linker comprises a position for inserting the nucleic acid

CC sequence and lacks the coding sequence of a gene with which it is

CC naturally associated. The vector is able to encode a hormone, a growth

CC factor, an enzyme, a clotting factor, an apolipoprotein, a receptor, a

CC drug or an antigen. The vectors can be used to treat wounds or surgical

CC incisions as well as diseases, such as psoriasis, cancer, alopecia and to

CC create transgenic animals for assessing human disease in an animal model.

CC The present sequence represents the loricrin gene

XX

SQ Sequence 6530 BP; 1635 A; 1507 C; 1696 G; 1692 T; 0 U; 0 Other;

Query Match 3.2%; Score 49.2; DB 2; Length 6530;

Best Local Similarity 46.1%; Pred. No. 0.0029;

Matches 165; Conservative 0; Mismatches 193; Indels 0; Gaps 0;

QY 96 GCCTCAGTACAGGATGCGCTCAGTACCGAGTCTCTCAGCGGACACCGTGCCTCTCT 155

Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

3530 GCCTCCGTAGTCTGGCACTGATACTGTTGGAGCAGTCCGCGCGCCACCGGAGGA 3471

QY 156 GCAACCCCATCATAGCCAGTCCGCGCTCCTCACTCGTACATGGGGCAGCGCGTACCG 215

Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

3470 ACCACCTCCGAGCTCGTGCCGCGCCACCGGAATAGCCGCGCGCGCAGTGA 3411

QY 216 GCCTGATCTGAACAGGTACCCCGCATCAAGTCACGATGTTTACGGTCTTCTGTCGCC 275

Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

3410 TCCACCGCCAGCTGAGTAGCCGCGCCACAGCTAGAGCCGCGCTCCAGAGAGCCGCC 3351

QY 276 GATAATGCCCCACACTACCGTGGGAGCTGCTCCGACATCTTTCTTCTCATCCAA 335

Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

3350 GGAATAGCCGCTCCGAGCTGGAGCCACCGCGCCACAGAGTACTTGCCGCTCCGGA 3291

QY 336 TCCGAGGCGGAGGACAGGCAATCGCGCACTATCCTCCTCATAGCGTCT 395

Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

3290 ACCGCGCGCAGCCAGCTGCGGCTCCAGAGAGCCTCCGAGTAGGAGCCGCC 3231

QY 396 CCGCGCGCTTCCAGCGCTCAGTCGTACCGGAGCCAAATTGCGCGCGCTCCCGG 453

Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

3230 GCCTCCTGATCCGCTCCGAGCTGGAGCGCGCGCGCGCGCGCGCGCGCGCGCG 3173

RESULT 12

AAC68948/c

ID AAC68948 standard; DNA; 6530 BP.

XX

AC AAC68948;

	Best Local Similarity Matches	54.9%; Conservative	Pred. No. 0.0031; 0; Mismatches	79; Indels	0; Gaps	0;
QY	695	CCAAGGATGCCGATGGAAGTTCCTTGCCCGAACTGTAAACAAGACTTATCTTCATGCCA				754
Db	86	CCGGGGTGGCGATGGCCCGTTCGCTGCCCACTCTGCTGGAAGGTTTTCAAGAAAGCCCA				145
QY	755	AGCATCTCAAGCGCCATCTGCTACGCCACACTGGTGACCGCCCGTACATGTGTGTTCTTT				814
Db	146	GTCACCTCCACCAGCACACAGATCATCCACACGGGGCGAGAGCCCTTCTCCTGCTCGGTGT				205
QY	815	GCAAAGACACCTTCTCTGCGAGTGATATCCTGAAAAGTCAATTTCCAAAAATGCTC				869
Db	206	GCAGCAAAAGCTTCAACCGCAGGGAGAGTCTGAAGCGCCACGTGAAGACGCACATC				260

RESULT 14
AAC93730
ID AAC93730 standard; cDNA; 549 BP.
XX
XX AAC93730;
AC
XX
DT 19-FEB-2001 (first éntry)
XX
DE Cat flea hindgut and Malpighian tubule (HMT) cDNA, SEQ ID NO:225.
XX
KW Cat flea; hindgut and Malpighian tubule nucleic acid; HMT;
KW flea infestation; vaccine; antiparasitic; therapeutic target; diagnosis;
KW detection; ss.

The invention relates to novel cat flea (*Ctenocephalides felis*) nucleic acids which are expressed in hindgut and Malpighian tubule (HMT) tissue or head and nerve cord (HNC) tissue. The invention also relates to the encoded proteins. The invention additionally encompasses expression constructs, recombinant viruses and recombinant cells comprising the nucleic acids of the invention, recombinant production of the proteins, antibodies against the proteins, a method of identifying inhibitors of the proteins, and compositions comprising the inhibitors for administration to an animal. The nucleic acids, and the proteins they encode may be used in the prevention, treatment, and diagnosis of diseases associated with flea infestations. For example, the nucleic acids may be used to produce an HMT or HNC protein according to standard recombinant DNA methodology by inserting the nucleic acids into a host cell and culturing the cell to express the protein. The HMT and HNC nucleic acids may also be used as DNA probes in diagnostic assays (e.g., PCR) to detect and quantitate the presence of cat flea or other homologous nucleic acid sequences in samples. They may also be used to study the expression and function of the proteins and their role in metabolism. The HMT and HNC proteins may be used as antigens in the production of specific antibodies, and in assays to identify modulators (agonists and antagonists) of HMT and/or HNC protein expression and activity. The anti-

CC	HMT/HNC protein antibodies and antagonists may also be used to
CC	downregulate protein expression and activity. The antibodies may also be
CC	used as diagnostic agents for detecting the presence of flea polypeptides
CC	in samples (e.g., by enzyme linked immunosorbent assay (ELISA)). The
CC	present sequence represents a cat flea HMT cDNA of the invention
XX	
SQ	Sequence 549 BP; 170 A; 111 C; 114 G; 150 T; 0 U; 4 Other;
	Query Match 3.0%; Score 46.6; DB 3; Length 549;
	Best Local Similarity 55.2%; Pred. No. 0.0054;
	Matches 91; Conservative 0; Mismatches 74; Indels 0; Gaps 0
QY	697 AAGGATGCCGATGGAAAGTTCCCTTGCCCGAACTGTAAACAAGACTTATCTTCATGCCAAG 756
Db	219 AAATCTGCTGATGACATACACCGATGCGAACTGTGCAATAAGACGTTTAAATTGCAACA 278
QY	757 CATCTCAAGCGCCATCTGCTACGCCACACTGGTGACCGCCCGTACATGTGTCTCTTGC 816
Db	279 AATCTTAAGGCTCATATGGTTATGCACAGGGCGAACGGCCCTATGTCTGTGCCATATGC 338
QY	817 AAAGACACCTTCTCTCGCAGTGATATCCTGAAACGTCATTTCCAA 861
Db	339 AGTAAAGCTTTTCCACGGAATTGCCACCTTGAACCGACACATGCAA 383

RESULT 15
ADA71938/c
ID ADA71938 standard; DNA; 2000 BP.
XX
AC ADA71938;
XX
DT 20-NOV-2003 (first entry)
XX
DE Rice gene, SEQ ID 5263.
XX
KW Plant; bacterial infection; fungal infection; viral infection; rice;
KW gene; ds.
XX
OS Oryza sativa.
XX
PN WO2003000898-A1.
XX
PD 03-JAN-2003.
XX
PF 22-JUN-2001; 2001WO-IB001105.
XX
PR 22-JUN-2001; 2001WO-IB001105.
XX
PA (SYGN) SYNGENTA PARTICIPATIONS AG.
XX
PI Chang H, Chen W, Cooper B, Glazebrook J, Goff SA, Hou Y;
PI Katagiri F, Quan S, Tao Y, Whitham S, Xie Z, Zhu T, Zou G;
XX
DR WPI; 2003-175290/17.
XX
PT Identifying at least one gene involved in plant resistance or response to
PT pathogenic infection for conferring resistance or tolerance to a plant to
PT bacterial, fungal or viral infection by determining or detecting plant
PT gene expression.
XX
PS Claim 27; SEQ ID NO 5263; 899pp; English.
XX
CC The present invention relates to a method (M1) for identifying genes
CC involved in plant resistance or response to pathogenic infection. M1
CC comprises identifying a gene whose expression is significantly altered in
CC the incompatible interaction of plant gene expression relative to
CC expression of the gene in an uninfected plant, in a mutant plant that
CC does not express a gene associated with response to pathogenic infection,
CC or in a corresponding incompatible or compatible interaction. (M1) is
CC useful for conferring resistance to resistance or tolerance to a plant to
CC bacterial, fungal or viral infection. The present sequence was used to
CC illustrate the invention.
XX

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 3, 2004, 05:22:55 ; Search time 3726 Seconds
(without alignments)
12478.641 Million cell updates/sec

Title: US-10-029-180-7
Perfect score: 1557
Sequence: 1 atggatcctagaacacatcc.....ccgttcagccgccagacaa 1557

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

- EST:*
- 1: em_estba:*
 - 2: em_esthum:*
 - 3: em_estin:*
 - 4: em_estmu:*
 - 5: em_estov:*
 - 6: em_estpl:*
 - 7: em_estro:*
 - 8: em_htc:*
 - 9: gb_est1:*
 - 10: gb_est2:*
 - 11: gb_htc:*
 - 12: gb_est3:*
 - 13: gb_est4:*
 - 14: gb_est5:*
 - 15: em_estfun:*
 - 16: em_estom:*
 - 17: em_gss_hum:*
 - 18: em_gss_inv:*
 - 19: em_gss_pln:*
 - 20: em_gss_vrt:*
 - 21: em_gss_fun:*
 - 22: em_gss_mam:*
 - 23: em_gss_mus:*
 - 24: em_gss_pro:*
 - 25: em_gss_rod:*
 - 26: em_gss_phg:*
 - 27: em_gss_vrl:*
 - 28: gb_gss1:*
 - 29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	380.4	24.4	442	9	AA785001	AA785001 g4b03a1.r
2	185	11.9	719	14	CF945290	CF945290 TrEST-A39
3	171.6	11.0	776	12	BM077105	BM077105 TrEST-A34
4	108.8	7.0	687	10	BF251517	BF251517 EST418778

5	82.8	5.3	782	28	AQ161501	AQ161501 mgxb00008A
6	82	5.3	585	28	AQ255041	AQ255041 mgxb00008J
7	56.2	3.6	807	13	BX464554	BX464554 BX464554
c 8	53.8	3.5	1448	28	CC220110	CC220110 CH261-92F
9	52.6	3.4	1201	9	AL515475	AL515475 AL515475
10	52.2	3.4	727	13	BX319543	BX319543 BX319543
c 11	52.2	3.4	1970	29	CG748837	CG748837 P042-4-E0
12	51.4	3.3	1137	28	CC315110	CC315110 TAM32-30A
c 13	51.2	3.3	1839	29	CG747711	CG747711 P041-2-E0
c 14	50.6	3.2	567	28	AZ895651	AZ895651 RPCI-24-2
c 15	50.6	3.2	1616	29	CG753270	CG753270 P048-2-C0
16	50.2	3.2	916	29	CC678630	CC678630 OG0BP81TV
17	50.2	3.2	929	29	CG276224	CG276224 OG0GU57TH
c 18	50.2	3.2	939	29	CG287073	CG287073 OG2CB21TV
c 19	50.2	3.2	945	29	CC678642	CC678642 OG0BP81TV
20	50	3.2	1201	13	BX457568	BX457568 BX457568
21	49.6	3.2	1053	14	CK028030	CK028030 AGENCOURT
22	49.6	3.2	1201	13	BX421672	BX421672 BX421672
c 23	49.4	3.2	802	29	CNS011T0	AL100686 Drosophil
c 24	49.2	3.2	786	28	AQ745021	AQ745021 HS_5507_A
c 25	49.2	3.2	897	29	AG060530	AG060530 Pan trogl
c 26	49.2	3.2	915	13	BQ952432	BQ952432 AGENCOURT
27	49.2	3.2	1262	12	BM560943	BM560943 AGENCOURT
c 28	49.2	3.2	1287	10	BF256624	BF256624 HVSMEF001
29	49	3.1	481	9	AA695157	AA695157 GM02329.5
30	49	3.1	759	13	BU314584	BU314584 603543708
31	48.8	3.1	834	28	BZ051898	BZ051898 jnr57e08.
32	48.8	3.1	1101	29	CNS000U8	AL051971 Drosophil
33	48.8	3.1	1201	9	AL514195	AL514195 AL514195
c 34	48.8	3.1	1642	29	CG754720	CG754720 P050-2-A0
c 35	48.6	3.1	458	10	BF063108	BF063108 7h83e05.x
36	48.6	3.1	648	12	BI227246	BI227246 602949793
c 37	48.6	3.1	1026	29	AG081192	AG081192 Pan trogl
c 38	48.6	3.1	1793	29	CG754612	CG754612 P050-1-D0
c 39	48.4	3.1	732	29	BX149729	BX149729 Danio rer
c 40	48.4	3.1	815	29	BX239144	BX239144 Danio rer
41	48.4	3.1	987	29	CNS00418	AL066537 Drosophil
c 42	48.4	3.1	1137	29	AG078502	AG078502 Pan trogl
43	48.2	3.1	786	28	BH018846	BH018846 L2011a.d
c 44	48.2	3.1	2152	28	CC283954	CC283954 CH261-181
c 45	48	3.1	785	29	AG081505	AG081505 Pan trogl

ALIGNMENTS

RESULT 1	AA785001	AA785001	442 bp	mRNA	linear	EST 29-JUL-1998
LOCUS	g4b03a1.r1	Aspergillus nidulans	24hr asexual development	and	vegetative cDNA	lambda zap library
DEFINITION	g4b03a1 5',	mRNA sequence.				
ACCESSION	AA785001					
VERSION	AA785001.1	GI:2845169				
KEYWORDS	EST.					
SOURCE	Emericella nidulans (anamorph: Aspergillus nidulans)					
ORGANISM	Emericella nidulans					
REFERENCE	1 (bases 1 to 442)					
AUTHORS	Kupfer,D., Gray,J., Hausner,J., Lai,H., Martin,W., Aramayo,R., Prade,R. and Roe,B.					
TITLE	An Aspergillus nidulans EST Database					
JOURNAL	Unpublished (1998)					
COMMENT	Contact: Bruce A. Roe, University of Oklahoma, broe@ou.edu Department of Chemistry and Biochemistry Advanced Center for Genome Technology, University of Oklahoma 620 Parrington Oval, Norman, OK 73019, USA Tel: 405 325 4912 Fax: 405 325 7762 Email: broe@ou.edu We anticipate the future release of the cDNA clones to the Fungal Genetics Stock Center					

VERSION BM077105.2 GI:38355949
KEYWORDS EST.
SOURCE Hypocrea jecorina (anamorph: Trichoderma reesei)
ORGANISM Hypocrea jecorina
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
Hypocreomycetidae; Hypocreales; Hypocreaceae; Hypocrea.
REFERENCE 1 (bases 1 to 776)
AUTHORS Chamberg, F.S., Bonaccorsi, E.D., Ferreira, A.J.S., Ramos, A.S.P.,
Ferreira, J.R.Jr., Abrahao-Neto, J., Farah, J.P.S. and El-Dorry, H.
TITLE Elucidation of the metabolic fate of glucose in the filamentous
fungus Trichoderma reesei using expressed sequence tag (EST)
analysis and cDNA microarrays
JOURNAL J. Biol. Chem. 277 (16), 13983-13988 (2002)
MEDLINE 21950703
PUBMED 11825887
COMMENT On Feb 5, 2002 this sequence version replaced gi:18498287.
Contact: El-Dorry, Hamza
Department of Biochemistry
Institute of Chemistry. University of Sao Paulo
Avenida Professor Lineu Prestes, 748, Sao Paulo, SP, 05508-900,
BRASIL
Tel: (55) 11-38183848
Fax: (55) 11-38183848
Email: dorry@iq.usp.br
PCR Primers
FORWARD: Universal M13 forward primer
BACKWARD: Universal M13 reverse primer
Plate: 36 row: F column: 4
Seq primer: M13 reverse primer
High quality sequence stop: 582
POLYA=No.
FEATURES
source
1..776
/organism="Hypocrea jecorina"
/mol_type="mRNA"
/strain="QM9414 (ATCC26921)"
/db_xref="taxon:51453"
/clone="Tr-A3424"
/sex="Asexual"
/tissue_type="Mycelia"
/dev_stage="18 hr Glycerol-grown culture"
/lab_host="E. coli SOLR cells (kanamycin resistant)"
/clone_lib="TrEST-A"
/note="Vector: pBluescript SK(+); Site_1: EcoRI; Site_2:
XhoI; anamorph=Trichoderma reesei; Cloned
unidirectionally, 5' end of the cDNA cloned into EcoRI
site of pBluescript. Primer: Oligo (dT). Average insert
size: 1,2 kb; Uni-ZAP XR Vector system -5' adaptor
sequence: 5'GAATTCGGCAGAG3' -3' adaptor sequence:
5'CTCGAGTTTCTTTTCTTTTCTTTT3'"

ORIGIN
Query Match 11.0%; Score 171.6; DB 12; Length 776;
Best Local Similarity 60.7%; Pred. No. 1.5e-29;
Matches 318; Conservative 0; Mismatches 199; Indels 7; Gaps 2;
QY 392 TGTCTCCCGCCGCTTCCAGCGCTCAGTCGTACCCGCGAGCAATTGCGCGGCTCCCC 451
Dd 207 TGATGCCCCAGGCCACCACTGCGGTCTCTCACCCCTCAGCGGATTGCTCCTGCGCGGTTG 266
QY 452 GGGACCGTGTGCTGACTTCAACAATGGACTTCCTTCAGGAGCATTCAGTTATTCGGACG 511
Dd 267 GCGGCGGGTCTCCTGTCTGCTGCTCCCAATGCCTCCCGGTGTCTCATGGCGCAACCCG 326
QY 512 GAAAGCTCAAGTTGGGACCCCGTTGCTGCGAATGGTGTGCGCCGTATCCCGGGAAGG 571
Dd 327 GTGTGGCTCGCCCTACGGCCCGGGTTCTCTGATGCAGCCCTAACCCGATCCTCCCGGACG 386
QY 572 ACTCCCCCGAACCCAGGTTGTGGTTCTCAGGGGCGACCGGGTATCCTTCGAGTGTTC 631
Dd 387 GCGAGAGCCGACCCACGTCGTGTGGTTCTCAGGGTCCCGAGGCATCTTGCCCGAGCGCTC 446
QY 632 CGGACGCGCAACTCCGGTCAACAATGGTGTAAACGGCACCGGCAAGAACACTACTATCC 691

Db 447 CTGGACGGCCTACT-----GCCAGGCTGCGGAACTGGCGCAAGAACACCGTCATCC 500
QY 692 CGGCCAAGGATCCGATGGAAAGTTCCCTTCCCGAACTGTAAACAGACTTATCTTCATG 751
Dd 501 CTGTCAAGGATCCGACGGCAATTCCTTCCCTCACTGCACAAAGACGTATCTGCATG 560
QY 752 CCAAGCATCTCAAGCGCCATCTGTCTACGCCACACTGGTGACCGCCGTACATGTGTGTTT 811
Dd 561 CCAAGCACCTGAAGCGCCACCTGTCTCGCCACACTGGTGATCGCCCTACATGTGTGTC 620
QY 812 TTTGCAAAAGACACCTTCTCTCGCAGTGATATCCTGAAACGTCA-TTCCAAAAATGCTCA 870
Dd 621 TCTGCCGGGATACCTTCTCCGAAGGACATCCTTCAGGGCTCACTTCCAAAAATGCTTC 680
QY 871 ATCAGGCGTGGTAACCCCAACCGGAGCAACGCACCTTGTGCGCACCC 914
Dd 681 ATCCGGCGGGCAACCCCAAGGCGCGCTTGCCCTTTTGGTCCC 724

RESULT 4
BF251517
LOCUS BF251517 687 bp mRNA linear EST 15-NOV-2001
DEFINITION EST418778 Coccidioides immitis spherule cDNA library Coccidioides
immitis cDNA clone CIAAH91 5' sequence, mRNA sequence.
ACCESSION BF251517
VERSION BF251517.1 GI:16931660
KEYWORDS EST.
SOURCE Coccidioides immitis
ORGANISM Coccidioides immitis
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
Onygenales; Mitosporic Onygenales; Coccidioides.
REFERENCE 1 (bases 1 to 687)
AUTHORS Gardner, M.J. and Kirkland, T.
TITLE Generation of ESTs from Coccidioides immitis spherule cDNA library
JOURNAL Unpublished (2000)
COMMENT Contact: Malcolm J. Gardner
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301 838 3519
Fax: 301 838 0208
Email: gardner@tigr.org
FEATURES
source
1..687
/organism="Coccidioides immitis"
/mol_type="mRNA"
/db_xref="taxon:5501"
/clone="CIAAH91"
/dev_stage="spherule"
/lab_host="SOLR"
/clone_lib="Coccidioides immitis spherule cDNA library"
/note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
XhoI"

ORIGIN
Query Match 7.0%; Score 108.8; DB 10; Length 687;
Best Local Similarity 56.2%; Pred. No. 9.3e-15;
Matches 396; Conservative 0; Mismatches 252; Indels 56; Gaps 8;
QY 95 TGCCTCAGTACAGTCAGCCTCAGTACCCAGTCTCTCAGCCGCACACCCCTGCTCCTC 154
Dd 30 TCCACCAATATCAGATGCATTCGCCGTACGGCATGGCGCAACCGCATACGTTGCCGCAT 89
QY 155 TGCAACCCCATCATAGCCAGTCGCCCGCTCCTCACTCGTACATGGGGCAGCCCGGTACC 214
Dd 90 TGCAACACCATCAGATCCGTGCCAAATTCAGCATACGTATCTC-----CAGCCTTTCC 143
QY 215 GGCCTGATCTGAACAGGTACCCCGCA---TCAAGTCACGATGTTTACGGCTTCTGTGCTG 271
Dd 144 GGAATGACATGCCTAGATACCCGACAAACCTCCGGCACCGATGTTTATGCGGTAFCGTGCG 203
QY 272 CGCCGATAATGCCCCACACTACCGTGGGAGCTTGCTCCGACATCTTTCTTCTCATC 331

```
Db 204 CGCGTTGACTACGATGCGCGCTCAACAGTTTGCCACCGTCAACTTTTCTT----- 256
QY 332 CCAATCCGAGGCGCAGGACAGGCGCAGCAATCGCGGACTATCTCTCTCTCATAGCG 391
Db 257 -----GGCCACCAACATCCACAGCAATTCAGCCACATCATATA 293
QY 392 TGTCTCCGCGCGCTTCCAGCGCTCAGTCAGTACCGGACGCAATTCGCGCGCTCTCCCG 451
Db 294 TGTCTCCACCCACGACAGCCACCAAGCATACCGGCAACCAATCGCGCCAGCCACCGC 353
QY 452 GGGACCGTCTGCTGACTTCAACAATGGACTTCCTTCAGGAGCATTCAGTTATTTCGACG 511
Db 354 GTGACAGAAGATCAGAGTATGCGCTG---TGCCGTTGSCCGCTTACAAACGGGGAGA 410
QY 512 GAAAGCCTCAAGGTTGGGACCCCGTTGCTG-----CGAATGGTGCTGCGCCGATCCCG 565
Db 411 ACAAGCAACCCGTATGGGCTGGCGCAGAGGACTGCCACCCACCTCTGGCCCCATATGC 470
QY 566 GGAAGGACTCCCCCGCAACCCAGGTTGTTGTTCTCAGGGGCGACGGGTATCTCTCCGA 625
Db 471 CAAAAGATCCTCCGAGGACACAAGTGGTAGGAACCCAGGCGGAGAGGTATTTCTCCGA 530
QY 626 GTGTTCCGGGACCGGCAACTCCGGTCACAAAATGTTAAACGGCACCGGCAAGAACTA 685
Db 531 GTGTACCGGGTGGCGTGC---GGTTACAAACGGTGTGAATGGCAC---CAAGGGAACAA 583
QY 686 CTATCCCGGCCAAGGATGCCGA-TGGAAAGTTCCCTTGCCCGCAACTGTAAACAAGACTTAT 744
Db 584 CCATCCCTGCCAAGGATGCAGACGGGAAAGTTTCCCTGTCCGCACTGCAATAAGACTTTA 643
QY 745 CTTCATGCCAAGCATCTCAAGCGCCATCTGCTACGCCACACTGG 788
Db 644 CTGGCATGCGAAGCACCTTAAGCGCATCTCTTACGCCACACCGG 687
```

```
RESULT 5
AQ161501
LOCUS
DEFINITION
  clone mgxb0008A02r, CUGI Rice Blast BAC Library Magnaporthe grisea genomic
  clone mgxb0008A02r, genomic survey sequence.
ACCESSION
  AQ161501
VERSION
  AQ161501.1 GI:3557902
KEYWORDS
  GSS.
SOURCE
  Magnaporthe grisea (anamorph: Pyricularia grisea)
  Magnaporthe grisea
  Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
  Sordariomycetes incertae sedis; Magnaporthaceae; Magnaporthe.
  1 (bases 1 to 782)
AUTHORS
  Yu,Y., Zhu,H., Boyd,C.A., Gaudette,B., Gayle,A., Kingsbury,R.,
  Phillips,K., Sasinowski,M, Wing,R.A. and Dean,R.A.
TITLE
  A BAC End Sequencing Framework to Sequence the Magnaporthe grisea
  Genome
JOURNAL
  Unpublished (1998)
COMMENT
  Contact: Dean RA
  Clemson University Genomics Institute
  Clemson University
  100 Jordan Hall, Clemson University, Clemson, SC 29634
  Tel: 864 656 5737
  Fax: 864 656 4293
  Email: rdean@clemson.edu
  Seq primer: GGAACACGCTATGACCATG
  Class: BAC ends
  High quality sequence stop: 315.
  Location/Qualifiers
    1..782
      /organism="Magnaporthe grisea"
      /mol_type="genomic DNA"
      /strain="70-15"
      /db_xref="taxon:148305"
      /clone="mgxb0008A02r"
      /tissue_type="Protoplasts"
      /lab_host="E. coli DH10B"
```

FEATURES

source

1..782

/organism="Magnaporthe grisea"

/mol_type="genomic DNA"

/strain="70-15"

/db_xref="taxon:148305"

/clone="mgxb0008A02r"

/tissue_type="Protoplasts"

/lab_host="E. coli DH10B"

ORIGIN

```
Query Match      5.3%; Score 82.8; DB 28; Length 782;
Best Local Similarity 51.5%; Pred. No. 1.3e-08;
Matches 271; Conservative 0; Mismatches 242; Indels 13; Gaps 3;

QY 321 CCTTTCTCATCCCAATCCGCGAGGCGCAGGCACAGGCGCAGCAATCGCCGCATATCCTCC 380
Db 116 CCCGTACCCCTGGCATGATGCCCGCGCGCGAGTACCAACAACTCACCCCGGCTATCCAC 175
QY 381 TCCTCATAGCGTGTCTCCGCGCGCTTCCAGCGCTCAGTCGTACCCGCGAGCCCAATTGCGCC 440
Db 176 CACAAACCTCAATGATGACCCAGGCATCGACCGCAATGGCCACCCCTCAACCTATTGCGCC 235
QY 441 GGCGCTCCCCGGGACCGTGTGTGCTGACTTCAACAATGGACTTCTTCAGGAGCATTCAG 500
Db 236 CGCGCTCTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 295
QY 501 TTATTTCGGACGGAAAG-----CCTCAAGGTTGGGACCCCGTTGCTGCGAATGGTGTGCG 555
Db 296 CCCTCAGGGCGGCATGCTCGCCCATACGCGCAGAGCCCATTTGATGCCCCAG--CCTCAA 353
QY 556 CCGTATCCCGGGAAGGACTCCCCCGAAACCCAGGTTGTTGGTTCTCAGGGCGACGCGGT 615
Db 354 ATGATGCAACAGGAAGGCGACGCGCGCACATGTGTTGGTCTCTCAGGTCGCCGAGGA 413
QY 616 ATCCTTCCGAGTGTTCGCGGACGCGCAACTCCGGTCACAAATGGTGTAAACGGCACCGGC 675
Db 414 ATCTTCCCGAGCGCGCGGTAGG-----CCCCAGGCCCCGCTCGCTGGCAGCCAGCGG 467
QY 676 AAGAACACTACTATCCCGGCCAAGGATGCCGATGGCAAGTTCCCTTGCCCGAACTGTAAC 735
Db 468 GCAAAGAACCCAGATCCCCCAGAGGATGCCGACGGGAAGTTCCCGTGCCCTCATTTGTAAC 527
QY 736 AAGACTTATCTCATGCCAAGCATCTCAAGCGCCCATCTGCTACGCCACACTGGTGACCGC 795
Db 528 AAGACGTACTCTGACGCCAAGCATCTCAAGAGGCATCTTCTGAGACGTAAAGTTTGGCCT 587
QY 796 CCGTACATGTGTGTTCTTTGCAAAAGACACCTTCTCTCGCAGTGATA 841
Db 588 GCTTTGCTGTTCTGTACGTTGAAAGCTGCTCAACTGACACATTAATA 633
```

RESULT 6

AQ255041

LOCUS

DEFINITION

AQ255041

ACCESSION

VERSION

AQ255041.1 GI:3779356

KEYWORDS

SOURCE

ORGANISM

Magnaporthe grisea (anamorph: Pyricularia grisea)

Magnaporthe grisea

Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;

Sordariomycetes incertae sedis; Magnaporthaceae; Magnaporthe.

1 (bases 1 to 585)

Yu,Y., Zhu,H., Boyd,C.A., Gaudette,B., Gayle,A., Kingsbury,R.,

Phillips,K., Sasinowski,M, Wing,R.A. and Dean,R.A.

A BAC End Sequencing Framework to Sequence the Magnaporthe grisea

Genome

Unpublished (1998)

JOURNAL

/clone_lib="CUGI Rice Blast BAC Library"

/note="Vector: pBACWICH; Site 1: HindIII; Site 2: HindIII; Rice blast is one of the most devastating fungal diseases of rice world wide. It is a filamentous ascomycete with a haploid genome (n=7) of approximately 40 Mbp. Rice blast is an important model fungal pathogen for studying numerous aspects of the fungal-host interaction. In order to facilitate genome wide analysis, a BAC library containing 9216 clones with an average insert size of 130 kbp was constructed. This library represents greater than 25X genome coverage. High density colony filters are available upon request."

Query Match 5.3%; Score 82.8; DB 28; Length 782;

Best Local Similarity 51.5%; Pred. No. 1.3e-08;

Matches 271; Conservative 0; Mismatches 242; Indels 13; Gaps 3;

QY 321 CCTTTCTCATCCCAATCCGCGAGGCGCAGGCACAGGCGCAGCAATCGCCGCATATCCTCC 380

Db 116 CCCGTACCCCTGGCATGATGCCCGCGCGCGAGTACCAACAACTCACCCCGGCTATCCAC 175

QY 381 TCCTCATAGCGTGTCTCCGCGCGCTTCCAGCGCTCAGTCGTACCCGCGAGCCCAATTGCGCC 440

Db 176 CACAAACCTCAATGATGACCCAGGCATCGACCGCAATGGCCACCCCTCAACCTATTGCGCC 235

QY 441 GGCGCTCCCCGGGACCGTGTGTGCTGACTTCAACAATGGACTTCTTCAGGAGCATTCAG 500

Db 236 CGCGCTCTCTGT 295

QY 501 TTATTTCGGACGGAAAG-----CCTCAAGGTTGGGACCCCGTTGCTGCGAATGGTGTGCG 555

Db 296 CCCTCAGGGCGGCATGCTCGCCCATACGCGCAGAGCCCATTTGATGCCCCAG--CCTCAA 353

QY 556 CCGTATCCCGGGAAGGACTCCCCCGAAACCCAGGTTGTTGGTTCTCAGGGCGACGCGGT 615

Db 354 ATGATGCAACAGGAAGGCGACGCGCGCACATGTGTTGGTCTCTCAGGTCGCCGAGGA 413

QY 616 ATCCTTCCGAGTGTTCGCGGACGCGCAACTCCGGTCACAAATGGTGTAAACGGCACCGGC 675

Db 414 ATCTTCCCGAGCGCGCGGTAGG-----CCCCAGGCCCCGCTCGCTGGCAGCCAGCGG 467

QY 676 AAGAACACTACTATCCCGGCCAAGGATGCCGATGGCAAGTTCCCTTGCCCGAACTGTAAC 735

Db 468 GCAAAGAACCCAGATCCCCCAGAGGATGCCGACGGGAAGTTCCCGTGCCCTCATTTGTAAC 527

QY 736 AAGACTTATCTCATGCCAAGCATCTCAAGCGCCCATCTGCTACGCCACACTGGTGACCGC 795

Db 528 AAGACGTACTCTGACGCCAAGCATCTCAAGAGGCATCTTCTGAGACGTAAAGTTTGGCCT 587

QY 796 CCGTACATGTGTGTTCTTTGCAAAAGACACCTTCTCTCGCAGTGATA 841

Db 588 GCTTTGCTGTTCTGTACGTTGAAAGCTGCTCAACTGACACATTAATA 633

AQ255041

mgxb0008J11r CUGI Rice Blast BAC Library Magnaporthe grisea genomic clone mgxb0008J11r, genomic survey sequence.

AQ255041

AQ255041.1 GI:3779356

GSS.

Magnaporthe grisea (anamorph: Pyricularia grisea)

Magnaporthe grisea

Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;

Sordariomycetes incertae sedis; Magnaporthaceae; Magnaporthe.

1 (bases 1 to 585)

Yu,Y., Zhu,H., Boyd,C.A., Gaudette,B., Gayle,A., Kingsbury,R.,

Phillips,K., Sasinowski,M, Wing,R.A. and Dean,R.A.

A BAC End Sequencing Framework to Sequence the Magnaporthe grisea

Genome

Unpublished (1998)

JOURNAL

COMMENT	Contact: Dean RA Clemson University Genomics Institute Clemson University 100 Jordan Hall, Clemson University, Clemson, SC 29634 Tel: 864 656 5737 Fax: 864 656 4293 Email: rdean@clemson.edu Seq primer: GGAACAGCTATGACCATG Class: BAC ends High quality sequence stop: 236.			DEFINITION BX464554 Homo sapiens PLACENTA Homo sapiens cDNA clone CS0DE007YB11 3-PRIME, mRNA sequence. BX464554 VERSION BX464554.1 GI:31021646 KEYWORDS EST. SOURCE Homo sapiens (human) ORGANISM Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 1 (bases 1 to 807) Li,W.B., Gruber,C., Jessee,J. and Polayes,D. Full-length cDNA libraries and normalization Unpublished (2001) Contact: Genoscope Genoscope - Centre National de Sequencage BP 191 91006 EVRY cedex - France Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr Library was constructed by Life Technologies, a division of Invitrogen. This sequence belongs to sequence cluster 3370.r For more information about this cluster, see http://www.genoscope.cns.fr/ cgi-bin/cluster.cgi?seq=CS1DE002ZF03NP1&cluster=3370.r. Contact : Feng Liang Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com/ InVitroGen Corporation 1600 Paraday Avenue Genoscope sequence ID : CS1DE002ZF03NP1.		
	Location/Qualifiers 1..807 /organism="Homo sapiens" /mol_type="mRNA" /db_xref="taxon:9606" /clone="CS0DE007YB11" /tissue_type="PLACENTA" /clone_lib="Homo sapiens PLACENTA" /note="Vector: pCMVSPORT 6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoRV sites of the pCMVSPORT 6 vector. Library was not normalized."			FEATURES source 1..807 /organism="Homo sapiens" /mol_type="mRNA" /db_xref="taxon:9606" /clone="CS0DE007YB11" /tissue_type="PLACENTA" /clone_lib="Homo sapiens PLACENTA" /note="Vector: pCMVSPORT 6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoRV sites of the pCMVSPORT 6 vector. Library was not normalized."		
ORIGIN	Query Match Best Local Similarity 53.2%; Score 82; DB 28; Length 585; Matches 248; Conservative 0; Mismatches 205; Indels 13; Gaps 3;			ORIGIN		
	321 CCTTCTCATCCCAATCCGAGGCGGACAGGCAGCAATCGCGCACTATCCTCC 380 116 CCGTACCTGGCATGATGCCGCGCGAGTACCAACCTACCCGGCTATCCAC 175 			Query Match Best Local Similarity 31.5%; Score 56.2; DB 13; Length 807; Matches 178; Conservative 97; Mismatches 290; Indels 0; Gaps 0;		
ORIGIN	381 TCCTCATAGTGCTCCCGCCGCTTCCAGGCTCAGTCGTACCAATGGACTTCCTTCAGGAGCATTCAG 440 176 CACAACCTCAATGATGACCCAGGCATCGACGCAATGGCCCCACCTCAACCTATTGCGCC 235 			Query Match Best Local Similarity 3.6%; Score 56.2; DB 13; Length 807; Matches 178; Conservative 97; Mismatches 290; Indels 0; Gaps 0;		
	441 GGCGCTCCCGGACCGTCGTGCTGACTTCAACAATGGACTTCCTTCAGGAGCATTCAG 500 236 CGCGCTCCTGGTGGTGGCTCCCGGCTTCCGCTCAATGCGTCCATGCCCCGGTGGCATGGT 295 			Query Match Best Local Similarity 3.6%; Score 56.2; DB 13; Length 807; Matches 178; Conservative 97; Mismatches 290; Indels 0; Gaps 0;		
ORIGIN	501 TTATTCGGACGAAAG-----CCTCAAGTTGGACCCCGTTGCTGCAATGGTGTGCG 555 296 CCTCAGGCGGCATGTGCTCGCCATACGGCGAGACCCATTGATGCCCCAG--CCTCAA 353 			Query Match Best Local Similarity 3.6%; Score 56.2; DB 13; Length 807; Matches 178; Conservative 97; Mismatches 290; Indels 0; Gaps 0;		
	556 CCGTATCCCGGAAGGACTCCCCCGAACCCAGGTTGTTGGTTCTCAGGGCGACGGGT 615 354 ATGATGCAACAGGAAGCGACGCGCCGACACATGTGTTGGTCTCAGGGTCCCGAGGA 413 			Query Match Best Local Similarity 3.6%; Score 56.2; DB 13; Length 807; Matches 178; Conservative 97; Mismatches 290; Indels 0; Gaps 0;		
ORIGIN	616 ATCCTTCGAGTGTCCGGGACGGCAACTCCGGTCACAAATGGTGTAAACGGCACCGGC 675 414 ATCCTCCGAGCGCGCGGTAGG-----CCCCAGGCCCGGTGCTGCGTGGCAGCGCG 467 			Query Match Best Local Similarity 3.6%; Score 56.2; DB 13; Length 807; Matches 178; Conservative 97; Mismatches 290; Indels 0; Gaps 0;		
	676 AAGAACAATACTATCCCGGCCAAGGATGCGGATGGAAGTTCCCTTGCCCGAACTGTAAC 735 468 GCAAAGAACCATATCCCCAGAGGATGCGACGGGAAGTTCCCGTGCCCTCATTTGTAAC 527 			Query Match Best Local Similarity 3.6%; Score 56.2; DB 13; Length 807; Matches 178; Conservative 97; Mismatches 290; Indels 0; Gaps 0;		
ORIGIN	736 AAGACTTATCTCATGCCAAGCATCTCAGCGCCCATCTGCTACGCC 781 528 AAGACGTACCTGCACGCCAAGCATCTCAAGAGGCATCTTCTGAGAC 573 			Query Match Best Local Similarity 3.6%; Score 56.2; DB 13; Length 807; Matches 178; Conservative 97; Mismatches 290; Indels 0; Gaps 0;		
	RESULT 7 BX464554 LOCUS			Query Match Best Local Similarity 3.6%; Score 56.2; DB 13; Length 807; Matches 178; Conservative 97; Mismatches 290; Indels 0; Gaps 0;		
ORIGIN	807 bp mRNA linear EST 22-MAY-2003			Query Match Best Local Similarity 3.6%; Score 56.2; DB 13; Length 807; Matches 178; Conservative 97; Mismatches 290; Indels 0; Gaps 0;		
	BX464554			Query Match Best Local Similarity 3.6%; Score 56.2; DB 13; Length 807; Matches 178; Conservative 97; Mismatches 290; Indels 0; Gaps 0;		


```
QY      260 CGTCTTCTGCTGGCCGATAATGCCCCACACTACCGTGGGAGCTTGCTCCGACATCTT 319
      ||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db      324 CCCCSSSCSCCCCCCCCCCCCCCCCCCTTCCCCSCCCCCSCCCCTCCCCGSCCC 383

QY      320 TCCTTCTCATCCCAATCCGAGGCGCAGGCACAGGCGCAGCAATCGCCGCACTATCCTC 379
      ||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db      384 TCSCTYCCCCCCCCCCCCCGCCCCCCCCCCCCCAACCCACACACACACACACCTCTTCCC 443

QY      380 CTCCTCATAGCGTGCTCCCGCGGCTTCCAGCGCTCAGTCGTCAGTACCCGCGCAATTGCGC 439
      ||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db      444 CGCGGSSSCCCCCCTCCMGSCCCCYCCCCCGSCCCCTCTCCCCCCCCCCCCCCCC 503
```

```
QY      440 CGGCGCCTCCCCGGGACCGTCGTGCTGACTTCAACAA 476
      ||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db      504 CGSSSCCCCTGTCMCCCCCGGCGSCCTCCACCA 540
```

```
RESULT 10
LOCUS   BX319543
DEFINITION
BX319543 AGENAE Rainbow trout multi-tissues subtracted library
(tcay) Oncorhynchus mykiss cDNA clone tcay0037b.k.13 5prim, mRNA
sequence.
ACCESSION
VERSION  BX319543.1  GI:29600188
KEYWORDS
SOURCE   EST.
ORGANISM Oncorhynchus mykiss (rainbow trout)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei;
Protacanthopterygii; Salmoniformes; Salmonidae; Oncorhynchus.
REFERENCE
AUTHORS  Govoroun,M., Guiguen,Y. and Le Gac,F.
TITLE    Construction and primary characterization of normalized cDNA
libraries in rainbow trout, Oncorhynchus mykiss
JOURNAL  Unpublished (2003)
COMMENT  Contact: Guiguen Y
INRA - SCRIBE
Campus de beaulieu, RENNES cedex, 35042, France
Tel: 02.23.48.50.09
Fax: 02.23.48.50.20
Email: Yann.Guiguen@beaulieu.rennes.inra.fr
Sequence cleaned of vector, adaptor and repetitions. Contact us
at sigenasupport@jouy.inra.fr to obtain the chromatogram of this
sequence.
Plate: 0037 row: k column: 13
Seq primer: M13F.
```

```
FEATURES
source
Location/Qualifiers
1..727
/organism="Oncorhynchus mykiss"
/mol_type="mRNA"
/db_xref="taxon:8022"
/clone="tcay0037b.k.13"
/tissue_type="adipose tissue, blood, brain,
differentiating gonads, gills, interrenal, intestine,
kidney, liver, muscle, ovary, pituitary, testis"
/dev_stage="from embryos to adults"
/lab_host="DH10B"
/clone_lib="AGENAE Rainbow trout multi-tissues subtracted
library (tcay)"
/note="vector: pT7T3D-pac; Rainbow trout multi-tissues ~
normalized + 1 subtraction (tcay); Clone distribution :
AGENAE Resource centre. Francois PIUMI,
Francois.Piumi@jouy.inra.fr, INRA, CEA Radiobiologie et
Etude du genome (LREG), Domaine de Vilvert, 78352,
Jouy-en-Josas cedex, FRANCE"
```

ORIGIN

```
Query Match      3.4%; Score 52.2; DB 13; Length 727;
Best Local Similarity 55.9%; Pred. No. 0.21;
Matches 99; Conservative 0; Mismatches 78; Indels 0; Gaps 0;
```

```
QY      715 TTCCCTTGCCCGAAGTGAACAAGACTTATCTTCAATGCCAAGCATCTCAAGCGCCATCTG 774
      ||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db      471 TACACATGCCCCGAGTGTACGCGTCTCTTCAACCAAGGAGCCATCTCAAGTCCCACATG 530

QY      775 CTAGCCACACTGGTGACGCGCCGTACATGTGTCTTTGCAAAAGACACCTTCTCTCGC 834
      ||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db      531 CGCCTCCACACGCGGTGAGCGCGGTACAAGTGCCAGCAATGTGACAAAGTCTTCAACCAC 590

QY      835 AGTGATATCTGAAACGTCATTTCCAAAAATGCTCAATCAGGCGTGGTAACCCACC 891
      ||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db      591 AACGTGAGTCTGAAAAAGCCACGTCACGCGCTACCAACGAGCCTTGGCTACGGGACC 647
```

```
RESULT 11
LOCUS   CG748837/c
DEFINITION
P042-4-E01.za Ppa EcoRI BAC Library Pristionchus pacificus genomic,
genomic survey sequence.
ACCESSION
VERSION  CG748837.1  GI:37969763
KEYWORDS
SOURCE   GSS.
ORGANISM Pristionchus pacificus
Pristionchus pacificus
Eukaryota; Metazoa; Nematoda; Chromadorea; Diplogasterida;
Neodiplogasteridae; Pristionchus.
REFERENCE
AUTHORS  Srinivasan,J., Sinz,W., Jesse,T., Wiggers-Perebolte,L., Jansen,K.,
Buntjer,J., van der Meulen,M. and Sommer,R.J.
TITLE    An integrated physical and genetic map of the nematode Pristionchus
pacificus
JOURNAL  Mol. Genet. Genomics 269 (5), 715-722 (2003)
MEDLINE  22835951
PUBMED  12884007
COMMENT  Contact: Sommer RJ
Evolutionary Biology
Max-Planck-Institute for Developmental Biology
Spemannstr. 37-39, Tuebingen D-72076, Germany
Tel: 00497071601371
Fax: 00497071601498
Email: ralf.sommer@tuebingen.mpg.de
Class: BAC ends.
```

FEATURES

```
Location/Qualifiers
1..1970
/organism="Pristionchus pacificus"
/mol_type="genomic DNA"
/strain="California"
/db_xref="taxon:54126"
/clone_lib="Ppa EcoRI BAC Library"
/note="The library was generated by a partial digest of
the genomic DNA with EcoRI and cloning into the BAC
vector."
```

ORIGIN

```
Query Match      3.4%; Score 52.2; DB 29; Length 1970;
Best Local Similarity 36.7%; Pred. No. 0.35;
Matches 165; Conservative 0; Mismatches 285; Indels 0; Gaps 0;

QY      15 CCATCCCTCTCGGCCTCCGTCCTACCACTGCTGCCTCAAGGATCGGCGCCTCTTCTCTGC 74
      ||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db      1015 CCCCCCCCCCNCCCNCCCNCCCNCCCNCCCNCCCNCCCNCCCNCCCNCCCNCCCNCCCN 956

QY      75 TCCCATCTCGAGCATGCCAATGCCTCAGTACACGATGCAGCCTCAGTACCCAGTCTCTCA 134
      ||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db      955 CCCCCCCCCCCCCCCCCCNCCCNCCCNCCCNCCCNCCCNCCCNCCCNCCCNCCCNCCCNCC 896

QY      135 GCGGCACACCTGCCTCCTCTGTGAAACCCCATCATAGCCAGTCGCCGCTCCTCACTCGTA 194
      ||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db      895 CNCCCCCCCCCNCCCNCCCNCCCNCCCNCCCNCCCNCCCNCCCNCCCNCCCNCCCNCC 836

QY      195 CATGGGGAGCCCGCGTACCGGCTGATCTGAACAGGTACCCCGCATCAAGTCACGATGT 254
      ||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db      835 CNNCNCNNCCNNCCCNCCCNCCCNCCCNCCCNCCCNCCCNCCCNCCCNCCCNCCCNCCCN 776
```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 1, 2004, 17:38:48 ; Search time 59 Seconds
(without alignments)
2485.460 Million cell updates/sec

Title: US-10-029-180-8
Perfect score: 2866
Sequence: 1 MDPNRNHPSPSTSLPQGSA.....NGLYLASTMSGDGTQPARQ 519

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_29Jan04:*
1: _geneseqp1980s:*
2: _geneseqp1990s:*
3: _geneseqp2000s:*
4: _geneseqp2001s:*
5: _geneseqp2002s:*
6: _geneseqp2003as:*
7: _geneseqp2003bs:*
8: _geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result		%		Query		Match		Length		DB		ID		Description	
No.	Score														
1	2866	100.0	519	5	ABP63087									Abp63087 FLO11 gen	
2	2866	100.0	600	5	ABP63143									Abp63143 Fungal ge	
3	2866	100.0	611	5	ABP63142									Abp63142 Fungal ge	
4	255.5	8.9	722	4	AAM39029									Aam39029 Human pol	
5	251.5	8.8	2703	4	ABB63299									Abb63299 Drosophil	
6	251.5	8.8	2703	5	ABG70019									Abg70019 Larval vi	
7	249	8.7	554	4	ABG15134									Abg15134 Novel hum	
8	240	8.4	723	7	ADB64092									Adb64092 Human pro	
9	240	8.4	2062	7	ADC31187									Adc31187 Human nov	
10	239	8.3	1401	4	ABG224037									Abg224037 Novel hum	
11	238.5	8.3	1142	7	ADC07968									Adc07968 Rice prot	
12	237	8.3	1192	4	ABB65293									Abb65293 Drosophil	
13	236.5	8.3	644	4	ABB70436									Abb70436 Drosophil	
14	233.5	8.1	1185	2	AAy33497									Aay33497 Human atr	
15	229.5	8.0	926	4	ABB65135									Abb65135 Drosophil	
16	223	7.8	420	5	ABB57247									Abb57247 Mouse isc	
17	211	7.4	1963	4	ABB62819									Abb62819 Drosophil	
18	210.5	7.3	505	4	AAB61923									Aab61923 Human ear	
19	210.5	7.3	543	2	AAR24022									Aar24022 Human pro	
20	210.5	7.3	543	2	AAW76985									Aaw76985 Human EGR	
21	210.5	7.3	543	3	AAy51116									Aay51116 Human EGR	
22	210.5	7.3	543	4	AAU09066									Aau09066 Human ZIF	
23	210.5	7.3	543	4	AAB76861									Aab76861 Human lun	
24	210.5	7.3	543	5	AAU85516									Aau85516 Clone #18	
25	210.5	7.3	543	6	ABU69488									Abu69488 Human lun	

26	210.5	7.3	543	6	ABU66390									Abu66390 Lung canc	
27	210.5	7.3	543	7	ADB75271									Adb75271 Prostate	
28	210.5	7.3	543	7	ADC37141									Adc37141 543 amino	
29	210.5	7.3	543	7	ADD48114									Add48114 Human Pro	
30	210.5	7.3	543	8	ADE71608									Ade71608 Human lun	
31	209.5	7.3	496	2	AAW76984									Aaw76984 Mouse Egr	
32	209.5	7.3	533	1	AAP93113									Aap93113 Egr-1. 3/	
33	209.5	7.3	533	2	AAR63129									Aar63129 Mouse Egr	
34	209.5	7.3	533	4	AAG78054									Aag78054 Human zin	
35	209.5	7.3	533	5	ABB57356									Abb57356 Mouse isc	
36	209	7.3	398	5	AAM49116									Aam49116 Mouse Sp1	
37	207.5	7.2	398	5	AAM49117									Aam49117 Human Sp1	
38	207.5	7.2	398	7	ADC31406									Adc31406 Human nov	
39	207.5	7.2	978	7	ADC31742									Adc31742 Human nov	
40	207.5	7.2	1386	7	ADE59291									Ades9291 Rat Prote	
41	207.5	7.2	1386	7	ADE59295									Ades9295 Rat Prote	
42	207	7.2	488	5	AAO20509									Aao20509 Protein o	
43	206	7.2	428	5	AAE25076									Aae25076 Mouse ost	
44	206	7.2	551	7	ADC31398									Adc31398 Human nov	
45	205.5	7.2	345	2	AAW52577									Aaw52577 Human tra	

ALIGNMENTS

RESULT 1
ABP63087
ID ABP63087 standard; protein; 519 AA.
XX
AC ABP63087;
XX
DT 22-OCT-2002 (first entry)
XX
DE FLO11 gene expression regulator An09.
XX
KW Fungal gene expression regulator; fungicide; gene therapy; An01; An05;
An09; An10; An13; An17; An20; An28; An34; At01-1; At01-2; At03; At05;
At07; At08; At11; At14; At16; At18; At19; At20; At22; At24; At27; At32;
PC05; PC06; PC07; PC08; PC09; PC10; PC18; PC23; PC24; PC25; PC33; PC34;
KW FLO11; fungal invasion; secondary metabolite; At279; At286; At291; At320;
At322; An1000; At167; At221; At233; At240; At274; Pcl000; Pcl001;
KW lovF; lovE; lovastatin; Pc804; acvA; penicillin; antifungal.
XX
OS Unidentified.
XX
PN WO200257456-A2.
XX
PD 25-JUL-2002.
XX
PF 24-DEC-2001; 2001WO-US049911.
XX
PR 22-DEC-2000; 2000US-0257431P.
XX
PA (MICR-) MICROBIA.
XX
PI Cali BM, Madden KT, Milne TG, Zhang L, Silva JC, Trueheart J;
PI Holtzman D, Sherman A;
XX
DR WPI; 2002-627368/67.
DR N-PSDB; ABQ94220.
XX
PT New isolated or recombinant gene, or purified protein, useful in
regulating fungal gene expression of FLO11, lovF, lovE or acvA for the
production of enzymes, secondary metabolites or other commercially and
medically useful products.
XX
PS Claim 57; Page 37; 71pp; English.
XX
CC The present invention relates to novel fungal gene expression regulators
(ABQ94217-ABQ94285 and ABP63084-ABP63152). An01, An05, An09, An10, An13,
An17, An20, An28, An34, At01-1, At01-2, At03, At05, At07, At08, At11,
At14, At16, At18, At19, At20, At22, At24, At27, PC05, PC06, PC07,
PC08, PC09, PC10, PC18, PC23, PC24, PC25, PC33 and PC34 are FLO11 gene

CC expression regulators. FLO11 is required for fungal invasion and its
CC expression is believed to be regulated by factors that also modulate
CC secondary metabolite production. At279, At286, At291, At320, At322,
CC An1000, At167, At221, At233, At239, At240, At274, Pc1000 and Pc1001 are
CC lovF gene expression regulators, and At501 and At574 are love gene
CC expression regulators. lovF and love are believed to be involved in the
CC production of the secondary metabolite lovastatin. Pc804 is an acvA gene
CC metabolite penicillin. The fungal gene expression regulators and their
CC coding sequences are useful in regulating or manipulating the expression
CC of fungal genes that are involved in the production of enzymes, secondary
CC metabolites and other commercially and medically useful products, in
CC order to achieve maximum benefit. The genes may also be used to identify
CC genes relevant to fungal invasion which may act as targets for the
CC development of antifungal drugs

XX Sequence 519 AA;

Query Match 100.0%; Score 2866; DB 5; Length 519;
Best Local Similarity 100.0%; Pred. No. 5.9e-193;
Matches 519; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MDPNRHPSRPPSTSLPQGSAPLPSAPISSMMPQYTMQPYPVSPHTLPLQPHHSQSP 60
Db |||||
1 MDPNRHPSRPPSTSLPQGSAPLPSAPISSMMPQYTMQPYPVSPHTLPLQPHHSQSP 60
QY 61 APHSYMGQPPYRPDLNRYPASSHDVYASSAAPIMPTTVGSLPPTSFLSHPNPQAQAQ 120
Db |||||
61 APHSYMGQPPYRPDLNRYPASSHDVYASSAAPIMPTTVGSLPPTSFLSHPNPQAQAQ 120
QY 121 QSPHYPPHSHVLPASSAQSYPOPIAPAPRRDRADFNGLPSGAFSYSDGKPGQWD PVA 180
Db |||||
121 QSPHYPPHSHVLPASSAQSYPOPIAPAPRRDRADFNGLPSGAFSYSDGKPGQWD PVA 180
QY 181 ANGAAPYPGKDSPTQVVGSGRRGILPSVPGRATPVTNGVNGTGKNTTIPAKDADGKFP 240
Db |||||
181 ANGAAPYPGKDSPTQVVGSGRRGILPSVPGRATPVTNGVNGTGKNTTIPAKDADGKFP 240
QY 241 CPNCNKTYLHAKHLKRHLRHTGDRPYMCVLCCKDTFSRSDILKRHFQKCSIRRGNTGAT 300
Db |||||
241 CPNCNKTYLHAKHLKRHLRHTGDRPYMCVLCCKDTFSRSDILKRHFQKCSIRRGNTGAT 300
QY 301 HLSHNAHVKRSQQAAANPVKPVQDEVSTVPPNGIPGITYGEGAVNGNGLAPARPGY 360
Db |||||
301 HLSHNAHVKRSQQAAANPVKPVQDEVSTVPPNGIPGITYGEGAVNGNGLAPARPGY 360
QY 361 ADHQTMGFPMSVNGMRGQPEDAPFGGRPHQGAPWPQAPKQSPYLVQPGADPSGHQLNI 420
Db |||||
361 ADHQTMGFPMSVNGMRGQPEDAPFGGRPHQGAPWPQAPKQSPYLVQPGADPSGHQLNI 420
QY 421 DRNIEQVKPVVQDPKRPVMPGHPGELDWTSMFQPAPEGYMFQSQMPGGQEPIHAH 480
Db |||||
421 DRNIEQVKPVVQDPKRPVMPGHPGELDWTSMFQPAPEGYMFQSQMPGGQEPIHAH 480
QY 481 VETERKYPTTTAGQESGMNGLYLASTWMSGDGT VQPARQ 519
Db |||||
481 VETERKYPTTTAGQESGMNGLYLASTWMSGDGT VQPARQ 519

RESULT 2

ABP63143
ID ABP63143 standard; protein; 600 AA.

XX ABP63143;

AC 22-OCT-2002 (first entry)

XX Fungal gene expression regulator SEQ ID 120.

XX Fungal gene expression regulator; fungicide; gene therapy; An01; An05;
KW An09; An10; An13; An17; An20; An28; An34; At01-1; At01-2; At03; At05;
KW At07; At08; At11; At14; At16; At18; At19; At20; At22; At24; At27; At32;
KW Pc05; Pc06; Pc07; Pc08; Pc09; Pc10; Pc18; Pc23; Pc24; Pc25; Pc33; Pc34;

KW FLO11; fungal invasion; secondary metabolite; At279; At286; At291; At320;
KW At322; An1000; At167; At221; At233; At239; At240; At274; Pc1000; Pc1001;
XX lovF; lovE; lovastatin; Pc804; acvA; penicillin; antifungal.

OS Unidentified.

XX WO200257456-A2.

XX 25-JUL-2002.

XX 24-DEC-2001; 2001WO-US049911.

XX 22-DEC-2000; 2000US-0257431P.

XX (MICR-) MICROBIA.

XX Cali BM, Madden KT, Milne TG, Zhang L, Silva JC, Trueheart J;

PI Holtzman D, Sherman A;

XX WPI; 2002-627368/67.

XX New isolated or recombinant gene, or purified protein, useful in
PT regulating fungal gene expression of FLO11, lovF, love or acvA for the
PT production of enzymes, secondary metabolites or other commercially and
PT medically useful products.

XX Claim 67; Page 67; 71pp; English.

CC The present invention relates to novel fungal gene expression regulators
CC (ABQ94217-ABQ94285 and ABP63084-ABP63152). An01, An05, An09, An10, An13,
CC An17, An20, An28, An34, At01-1, At01-2, At03, At05, At07, At08, At11,
CC At14, At16, At18, At19, At20, At22, At24, At27, At32, Pc05, Pc06, Pc07,
CC Pc08, Pc09, Pc10, Pc18, Pc23, Pc24, Pc25, Pc33 and Pc34 are FLO11 gene
CC expression regulators. FLO11 is required for fungal invasion and its
CC expression is believed to be regulated by factors that also modulate
CC secondary metabolite production. At279, At286, At291, At320, At322,
CC An1000, At167, At221, At233, At239, At240, At274, Pc1000 and Pc1001 are
CC lovF gene expression regulators, and At501 and At574 are love gene
CC expression regulators. lovF and love are believed to be involved in the
CC production of the secondary metabolite lovastatin. Pc804 is an acvA gene
CC expression regulator. acvA is involved in the production of the secondary
CC metabolite penicillin. The fungal gene expression regulators and their
CC coding sequences are useful in regulating or manipulating the expression
CC of fungal genes that are involved in the production of enzymes, secondary
CC metabolites and other commercially and medically useful products, in
CC order to achieve maximum benefit. The genes may also be used to identify
CC genes relevant to fungal invasion which may act as targets for the
CC development of antifungal drugs

XX Sequence 600 AA;

Query Match 100.0%; Score 2866; DB 5; Length 600;
Best Local Similarity 100.0%; Pred. No. 7e-193;
Matches 519; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MDPNRHPSRPPSTSLPQGSAPLPSAPISSMMPQYTMQPYPVSPHTLPLQPHHSQSP 60
Db |||||
82 MDPNRHPSRPPSTSLPQGSAPLPSAPISSMMPQYTMQPYPVSPHTLPLQPHHSQSP 141

QY 61 APHSYMGQPPYRPDLNRYPASSHDVYASSAAPIMPTTVGSLPPTSFLSHPNPQAQAQ 120
Db |||||

Db 142 APHSYMGQPPYRPDLNRYPASSHDVYASSAAPIMPTTVGSLPPTSFLSHPNPQAQAQ 201
|||

QY 121 QSPHYPPHSHVLPASSAQSYPOPIAPAPRRDRADFNGLPSGAFSYSDGKPGQWD PVA 180
Db |||||

Db 202 QSPHYPPHSHVLPASSAQSYPOPIAPAPRRDRADFNGLPSGAFSYSDGKPGQWD PVA 261
|||

QY 181 ANGAAPYPGKDSPTQVVGSGRRGILPSVPGRATPVTNGVNGTGKNTTIPAKDADGKFP 240
Db |||||

Db 262 ANGAAPYPGKDSPTQVVGSGRRGILPSVPGRATPVTNGVNGTGKNTTIPAKDADGKFP 321
|||

QY 241 CPNCNKTYLHAKHLKRHLRHTGDRPYMCVLCCKDTFSRSDILKRHFQKCSIRRGNTGAT 300
|||

Db 322 CPNCNKTYLHAKHLKRHLRLRHTGDRPYMCVLCCKDTFSRSDILKRHFQKCSIRRGNTGAT 381
QY 301 HLSHPNAHVKRSQQAAANPVKPVQDEVSSITVPPNGIPGTTYGEGAVNGNGLAPARPGY 360
Db 382 HLSHPNAHVKRSQQAAANPVKPVQDEVSSITVPPNGIPGTTYGEGAVNGNGLAPARPGY 441
QY 361 ADHQTMGFFPMSSVNGMGRGQPEDAFPGGRPHQGWAPQAPKQSPYLVQPGADPSGHQLNI 420
Db 442 ADHQTMGFFPMSSVNGMGRGQPEDAFPGGRPHQGWAPQAPKQSPYLVQPGADPSGHQLNI 501
QY 421 DRNIEQVKQPVVQDPKRPVMPGHPGHPGELDWTSMFQPAPEGYMFQSQMPGGQEPPIHAH 480
Db 502 DRNIEQVKQPVVQDPKRPVMPGHPGHPGELDWTSMFQPAPEGYMFQSQMPGGQEPPIHAH 561
QY 481 VETERKYYPPTTTAGQESGMNGLYLASTMSGDGTVPQPARQ 519
Db 562 VETERKYYPPTTTAGQESGMNGLYLASTMSGDGTVPQPARQ 600

RESULT 3
ABP63142
ID ABP63142 standard; protein; 611 AA.
XX
AC ABP63142;
XX
DT 22-OCT-2002 (first entry)
XX
DE Fungal gene expression regulator SEQ ID 119.
XX
KW Fungal gene expression regulator; fungicide; gene therapy; An01; An05;
KW An09; An10; An13; An17; An20; An28; An34; At01-1; At01-2; At03; At05;
KW At07; At08; At11; At14; At16; At18; At19; At20; At22; At24; At27; At32;
KW Pc05; Pc06; Pc07; Pc08; Pc09; Pc10; Pc18; Pc23; Pc24; Pc25; Pc33; Pc34;
KW FLO11; fungal invasion; secondary metabolite; At279; At286; At291; At320;
KW At322; An1000; At167; At221; At233; At239; At240; At274; Pc1000; Pc1001;
KW lovF; lovE; lovastatin; Pc804; acvA; penicillin; antifungal.
XX
OS Unidentified.
XX
XX WO200257456-A2.
PN
XX
PD 25-JUL-2002.
XX
PF 24-DEC-2001; 2001WO-US049911.
XX
PR 22-DEC-2000; 2000US-0257431P.
XX
PA (MICR-) MICROBIA.
XX
XX Cali BM, Madden KT, Milne TG, Zhang L, Silva JC, Trueheart J;
PI Holtzman D, Sherman A;
XX
DR WPI; 2002-627368/67.
XX
XX New isolated or recombinant gene, or purified protein, useful in
PT regulating fungal gene expression of FLO11, lovF, lovE or acvA for the
PT production of enzymes, secondary metabolites or other commercially and
PT medically useful products.
XX
PS Claim 67; Page 67; 71pp; English.
XX

CC The present invention relates to novel fungal gene expression regulators
CC (ABQ94217-ABQ94285 and ABP63084-ABP63152). An01, An05, An09, An10, An13,
CC An17, An20, An28, An34, At01-1, At01-2, At03, At05, At07, At08, At11,
CC At14, At16, At18, At19, At20, At22, At24, At27, At32, Pc05, Pc06, Pc07,
CC Pc08, Pc09, Pc10, Pc18, Pc23, Pc24, Pc25, Pc33 and Pc34 are FLO11 gene
CC expression regulators. FLO11 is required for fungal invasion and its
CC expression is believed to be regulated by factors that also modulate
CC secondary metabolite production. At279, At286, At291, At320, At322,
CC An1000, At167, At221, At233, At239, At240, At274, Pc1000 and Pc1001 are
CC lovF gene expression regulators, and At501 and At574 are lovE gene
CC expression regulators. lovF and lovE are believed to be involved in the
CC production of the secondary metabolite lovastatin. Pc804 is an acvA gene

CC expression regulator. acvA is involved in the production of the secondary
CC metabolite penicillin. The fungal gene expression regulators and their
CC coding sequences are useful in regulating or manipulating the expression
CC of fungal genes that are involved in the production of enzymes, secondary
CC metabolites and other commercially and medically useful products, in
CC order to achieve maximum benefit. The genes may also be used to identify
CC genes relevant to fungal invasion which may act as targets for the
CC development of antifungal drugs
XX
SQ Sequence 611 AA;

Query Match 100.0%; Score 2866; DB 5; Length 611;
Best Local Similarity 100.0%; Pred. No. 7.1e-193;
Matches 519; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MDPRNHPSRPPSTSLPQGSAPLPSAPISSMPMPQYTMQPOYPVSQPHTLPLPQPHHSQSP 60
Db 93 MDPRNHPSRPPSTSLPQGSAPLPSAPISSMPMPQYTMQPOYPVSQPHTLPLPQPHHSQSP 152
QY 61 APHSYMGQPPYRPLNRYPASSHDVYASSAAPIMPHTTVGSLLPPTSFLSHPNPQAQAAQ 120
Db 153 APHSYMGQPPYRPLNRYPASSHDVYASSAAPIMPHTTVGSLLPPTSFLSHPNPQAQAAQ 212
QY 121 QSPHYPPPHSVLPASSAQSYPOPIAPAPRRDRADFNNGLP SGAFSYSDGKPGQWDPVA 180
Db 213 QSPHYPPPHSVLPASSAQSYPOPIAPAPRRDRADFNNGLP SGAFSYSDGKPGQWDPVA 272
QY 181 ANGAAPYPGKDSPTQTVVGSQRRGILPSVPGRATPVTVNGVNGTGKNTTIPAKDADGKFP 240
Db 273 ANGAAPYPGKDSPTQTVVGSQRRGILPSVPGRATPVTVNGVNGTGKNTTIPAKDADGKFP 332
QY 241 CPNCNKTYLHAKHLKRHLRLRHTGDRPYMCVLCCKDTFSRSDILKRHFQKCSIRRGNTGAT 300
Db 333 CPNCNKTYLHAKHLKRHLRLRHTGDRPYMCVLCCKDTFSRSDILKRHFQKCSIRRGNTGAT 392
QY 301 HLSHPNAHVKRSQQAAANPVKPVQDEVSSITVPPNGIPGTTYGEGAVNGNGLAPARPGY 360
Db 393 HLSHPNAHVKRSQQAAANPVKPVQDEVSSITVPPNGIPGTTYGEGAVNGNGLAPARPGY 452
QY 361 ADHQTMGFFPMSSVNGMGRGQPEDAFPGGRPHQGWAPQAPKQSPYLVQPGADPSGHQLNI 420
Db 453 ADHQTMGFFPMSSVNGMGRGQPEDAFPGGRPHQGWAPQAPKQSPYLVQPGADPSGHQLNI 512
QY 421 DRNIEQVKQPVVQDPKRPVMPGHPGHPGELDWTSMFQPAPEGYMFQSQMPGGQEPPIHAH 480
Db 513 DRNIEQVKQPVVQDPKRPVMPGHPGHPGELDWTSMFQPAPEGYMFQSQMPGGQEPPIHAH 572
QY 481 VETERKYYPPTTTAGQESGMNGLYLASTMSGDGTVPQPARQ 519
Db 573 VETERKYYPPTTTAGQESGMNGLYLASTMSGDGTVPQPARQ 611

RESULT 4
AAM39029
ID AAM39029 standard; protein; 722 AA.
XX
AC AAM39029;
XX
DT 22-OCT-2001 (first entry)
XX
DE Human polypeptide SEQ ID NO 2174.
XX
KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia.
XX
OS Homo sapiens.
XX
PN WO200153312-A1.
XX

PD 26-JUL-2001.
XX 26-DEC-2000; 2000WO-US034263.
PF 23-DEC-1999; 99US-00471275.
XX 21-JAN-2000; 2000US-00488725.
PR 25-APR-2000; 2000US-00552317.
PR 20-JUN-2000; 2000US-00598042.
PR 19-JUL-2000; 2000US-00620312.
PR 03-AUG-2000; 2000US-00653450.
PR 14-SEP-2000; 2000US-00662191.
PR 19-OCT-2000; 2000US-00693036.
PR 29-NOV-2000; 2000US-00727344.
XX (HYSE-) HYSEQ INC.
XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J, Zhao QA;
PI Zhou P, Goodrich R, Drmanac RT;
XX WPI; 2001-442253/47.
DR N-PSDB; AAI58185.
XX Novel nucleic acids and polypeptides, useful for treating disorders such
PT as central nervous system injuries.
XX Example 4; SEQ ID NO 2174; 10078pp; English.
XX The invention relates to human nucleic acids (AAI57798-AAI61369) and the
CC encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: Immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders. Note: The sequence data for this patent did not form
CC part of the printed specification
XX Sequence 722 AA;

Query Match 8.9%; Score 255.5; DB 4; Length 722;
Best Local Similarity 24.3%; Pred. No. 1.7e-09;
Matches 145; Conservative 46; Mismatches 156; Indels 249; Gaps 35;
QY 7 PSRPPSTSLPQGSAPLPSAPISSMPMPQYTMQDPYVVSQPHLPLPQLPHHSQSPAP-HSY 65
Db 215 PERPLAAT---SAAPP-----QPQPQPQDE-PRSVQPEPEPEPEPEATPTAPAA 264
QY 66 MGQPPYRPD-----LNRYPAS-----SH-DVY 86
Db 265 PEEPPAPPEFRQCVCQSFTQSWFLKGMRKHKASFDHACVCGRCFKEPWFLKNHMKVH 324
QY 87 ASSAAPIMPHTTVGSLLPPTSFLSHPNP-QAQAQAQSPH-----YPPHSHVLPASSAQ 140
Db 325 ASKLG-----LRAPGASGPAPAPQPPDLGLLAYEP-----LGPALL--- 362
QY 141 YPQPIAPAP-PRDRRADFNGLPSGAFSY-----SDGKPGQWDPVAANGAAPPYKGD--- 191
Db 363 ----LAPAPTAPERRE-----PPSLLGLSLRAGEGRPNG-----EGAEPPGGRSFGG 406
QY 192 -SPRTQVVGSGRR-----GILPSV-----PGR----- 213
Db 407 FRPLSSALPARARRHRAEPEEEVEEVAQEETWARGSLGSLASHRPRGEGPGHSASA 466
QY 214 -----ATPVTNG--VNGT-----GKNTTIPAKDADGKFFPCPNCKNTYLLHAKHLKRHL 259
Db 467 AGAQAARSTATQEENGLLVGGTRPEGGRGAT--GKD-----CPFCGKFSRAHHLKVHLR 518

QY 260 RHTGDRPYMVCVLCKDTFSRSDILKRHFQKCSIRRGNPTGATHLSHPNAHVKRS---QQQA 316
Db 519 VHTGERPYKCPCHCDYAGTQSGSLKYHLQRHREQRSAGPGPPPEPPPSQSGAPQSGA 578
QY 317 AANPVKPVQDEVSTVPPNGIPGTTYGEGAVNGNGLAPARPGYADHQTMGFPMSSVNGM 376
Db 579 KPSPQPATWVEGASSPRPPS-----SGAGPGSRRKPAASPG-----RULR-----NGR 620
QY 377 GRGQPE-----DAFPGGRPHQ-----GAPW-----P 397
Db 621 G-GEAEPLDLSLRAGPGGEAGPGGALHRCILFCPFATGAPELMALHLQVHHSRRARRRPP 679
QY 398 QAPKQSPYLVQPGADPSGHQNLNIDRNIEQVKQPVVQDPKRPVMPGHPG-----HPGE 449
Db 680 QADASPPY----ARVPSG-----ETPPSPSQEGEGSGLSRPGE 714
RESULT 5
ABB63299
ID ABB63299 standard; protein; 2703 AA.
XX AC ABB63299;
XX DT 26-MAR-2002 (first entry)
XX DE Drosophila melanogaster polypeptide SEQ ID NO 16689.
XX KW Drosophila; developmental biology; cell signalling; insecticide;
XX KW pharmaceutical.
XX OS Drosophila melanogaster.
XX PN WO200171042-A2.
XX PD 27-SEP-2001.
XX PF 23-MAR-2001; 2001WO-US009231.
XX PR 23-MAR-2000; 2000US-0191637P.
PR 11-JUL-2000; 2000US-00614150.
XX PA (PEKE) PE CORP NY.
XX PI Venter JC, Adams M, Li PWD, Myers EW;
XX WPI; 2001-656860/75.
DR N-PSDB; ABL07402.
XX PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signaling and cell-cell
PT interactions.
XX PS Disclosure; SEQ ID NO 16689; 21pp + Sequence Listing; English.
XX CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
CC ABB72072). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX SQ Sequence 2703 AA;

Query Match 8.8%; Score 251.5; DB 4; Length 2703;
Best Local Similarity 23.9%; Pred. No. 1.5e-08;
Matches 152; Conservative 49; Mismatches 208; Indels 227; Gaps 36;
QY 3 PRNHPSR----PPSTSLPQGSAPLPSA-----PISSMPMPQYTMQPYVVSQ----- 45

Db 292 PQLGPSQOYRTPTNTSRGQSPYPPAHGQNSGYPSSPQQQQQQQQQQQQAGQQPGGP 351
QY 46 -PHTLPP---LQPHHSQSPAPHSYMGQPPYRDLNRYPA--SSH-DVYASSAAPIMPHTT 98
Db 352 VPGGPPGTGQQPPQONTPTTSQYSPYQRYPTTPGLPAGGSNHRHTAYSTHOYPEPNRPW 411
QY 99 VGSLLPPTSFLSHPNPQAQAQSQSPHY-----PPPHSVL--PPASSAQSY-----P 142
Db 412 PGGSSPSPGSGHPLPPA-----SPHHVPLPQQPPPPPHVSAGGPPSPSGHAPSPSP 465
QY 143 QP--IAPAPRRDRADFNGLPSGAFYSYDGKPGQWDPVAANGAAPYPGKDSPTQVV-- 198
Db 466 QPSQASPSPHQELIGQNSDSSG-----GAHSGM-----GSGP-PGTFNPQ-QVMRP 511
QY 199 ----GSQRRGILPSVPGR---ATPVTNGVNGTGKNTTIPAKDADGKFCPCNCNTYLH 250
Db 512 TPSPTGSSGSRSMSPAVAQNHPISRPASN-QSSSGGPMQPPVAGGPPMP----- 562
QY 251 AKHLKRHLRHTGDRPYMCVLCCKDTFSRSDILKRHFQKCSIRRGNTGATHLSHPNAHVK 310
Db 563 -----PH-----PGMPGGP-----PQQQ 575
QY 311 RSQQQAAAN-----PVKPVQDEVSTVPPNGIPGTTY-----GEG 346
Db 576 QSQQQASNSASSASNSPQQTPTPPAPPNQGMMNMATPPPPQGAAGGYPMPPHMHGGY 635
QY 347 AVNGNGLAPARPGYADHQTMGF-----PMSSVNGMG--RGQPE 382
Db 636 KMGPGQSPGAQGYPPQPPQYPPGNYPPRPQYPPGAYATGPPPPPTSQAGAGGANSMP 695
QY 383 DAFPGGRPHQGAP-----W-PQAPKQSPYLVQPGADPSGHQNLIDRNIEQVK--- 428
Db 696 GAQAGGYPRGMPNHTGQYPPYQWVPSPQQT----VPGGAPGGAMVG---NHVQKGT 748
QY 429 -QPVVQDPKRPVMPGHP-----GHPGELDWTSMFQQAPEGY-----MFSQ 468
Db 749 PPPVVGPPPPQCGSPRPLNYLKQHLQHKGGYGG-----SPTPPQGPQGYNGPTGMHP 803
QY 469 SMPGGQEPIH---AHVETERKYYPTTTAGQESGMNG 501
Db 804 GMPMG-PPHHMGPPHGPNTMG-PPTSTPPQSQMLQG 837

RESULT 6

ABG70019
ID ABG70019 standard; protein; 2703 AA.
XX AC ABG70019;
AC ABG70019;
DT 05-NOV-2002 (first entry)
XX DE Larval viability associated protein #18.
XX KW Fruit fly; larval viability; insecticidal activity; maize; wheat; oat;
KW rye; sorghum; rice; barley; millet; turf; cotton; sugarcane; sugar beet;
KW oilseed rape; soybean; vegetable crop; fruit.

XX Os Drosophila melanogaster.

XX WO200257455-A2.

XX PD 25-JUL-2002.

XX PF 18-JAN-2002; 2002WO-US001568.

XX PR 18-JAN-2001; 2001US-0262351P.

XX PA (SYGN) SYNGENTA PARTICIPATIONS AG.

XX PI Stam L, Bachmann J, Broadus J, Kamdar KP;

XX DR WPI; 2002-590746/63.

DR N-PSDB; ABS51395.

XX Identifying inhibitors of activity of proteins essential for Drosophila larval viability comprises expressing in a host a protein essential for larval activity and identifying compounds that inhibit or interact with the protein.

XX Claim 1; Page 108-117; 169pp; English.

PS The invention describes a method of identifying compounds that inhibit the activity of, or that interact with a protein essential for Drosophila larval viability comprising expressing in a recombinant host a DNA molecule to produce a protein essential for larval viability. The method is useful for identifying compounds with insecticidal activity. Compounds identified are useful as insecticides in crops such as maize, wheat, oats, rye, sorghum, rice, barley, millet, turf, cotton, sugarcane, sugar beet, oilseed rape, soybeans, vegetable crops and fruits. This is the amino acid sequence of a fruit fly larval viability associated protein

XX Sequence 2703 AA;

Query Match 8.8%; Score 251.5; DB 5; Length 2703;

Best Local Similarity 23.9%; Pred. No. 1.5e-08;

Matches 152; Conservative 49; Mismatches 208; Indels 227; Gaps 36;

QY 3 PRNHPSR---PPSTSLPQGSAPLPSA-----PISSMPMPQYTMQPYVSO----- 45

Db 292 PQLGPSQOYRTPTNTSRGQSPYPPAHGQNSGYPSSPQQQQQQQQQQAGQQPGGP 351

QY 46 -PHTLPP---LQPHHSQSPAPHSYMGQPPYRDLNRYPA--SSH-DVYASSAAPIMPHTT 98

Db 352 VPGGPPGTGQQPPQONTPTTSQYSPYQRYPTTPGLPAGGSNHRHTAYSTHOYPEPNRPW 411

QY 99 VGSLLPPTSFLSHPNPQAQAQSQSPHY-----PPPHSVL--PPASSAQSY-----P 142

Db 412 PGGSSPSPGSGHPLPPA-----SPHHVPLPQQPPPPPHVSAGGPPSPSGHAPSPSP 465

QY 143 QP--IAPAPRRDRADFNGLPSGAFYSYDGKPGQWDPVAANGAAPYPGKDSPTQVV-- 198

Db 466 QPSQASPSPHQELIGQNSDSSG-----GAHSGM-----GSGP-PGTFNPQ-QVMRP 511

QY 199 ----GSQRRGILPSVPGR---ATPVTNGVNGTGKNTTIPAKDADGKFCPCNCNTYLH 250

Db 512 TPSPTGSSGSRSMSPAVAQNHPISRPASN-QSSSGGPMQPPVAGGPPMP----- 562

QY 251 AKHLKRHLRHTGDRPYMCVLCCKDTFSRSDILKRHFQKCSIRRGNTGATHLSHPNAHVK 310

Db 563 -----PH-----PGMPGGP-----PQQQ 575

QY 311 RSQQQAAAN-----PVKPVQDEVSTVPPNGIPGTTY-----GEG 346

Db 576 QSQQQASNSASSASNSPQQTPTPPAPPNQGMMNMATPPPPQGAAGGYPMPPHMHGGY 635

QY 347 AVNGNGLAPARPGYADHQTMGF-----PMSSVNGMG--RGQPE 382

Db 636 KMGPGQSPGAQGYPPQPPQYPPGNYPPRPQYPPGAYATGPPPPPTSQAGAGGANSMP 695

QY 383 DAFPGGRPHQGAP-----W-PQAPKQSPYLVQPGADPSGHQNLIDRNIEQVK--- 428

Db 696 GAQAGGYPRGMPNHTGQYPPYQWVPSPQQT----VPGGAPGGAMVG---NHVQKGT 748

QY 429 -QPVVQDPKRPVMPGHP-----GHPGELDWTSMFQQAPEGY-----MFSQ 468

Db 749 PPPVVGPPPPQCGSPRPLNYLKQHLQHKGGYGG-----SPTPPQGPQGYNGPTGMHP 803

QY 469 SMPGGQEPIH---AHVETERKYYPTTTAGQESGMNG 501

Db 804 GMPMG-PPHHMGPPHGPNTMG-PPTSTPPQSQMLQG 837

RESULT 7

ABG15134
ID ABG15134 standard; protein; 554 AA.

XX ABG15134;
XX 18-FEB-2002 (first entry)
XX Novel human diagnostic protein #15125.
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder.
XX Homo sapiens.
XX WO200175067-A2.
XX 11-OCT-2001.
XX 30-MAR-2001; 2001WO-US008631.
XX 31-MAR-2000; 2000US-00540217.
XX 23-AUG-2000; 2000US-00649167.
XX (HYSE-) HYSEQ INC.
XX Drmanac RT, Liu C, Tang YT;
XX WPI; 2001-639362/73.
XX N-PSDB; AAS79321.
XX New isolated polynucleotide and encoded polypeptides, useful in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits and to assess
XX biodiversity.
XX Claim 20; SEQ ID NO 45493; 103pp; English.
XX The invention relates to isolated polynucleotide (I) and polypeptide (II)
XX sequences. (I) is useful as hybridisation probes, polymerase chain
XX reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
XX and in recombinant production of (II). The polynucleotides are also used
XX in diagnostics as expressed sequence tags for identifying expressed
XX genes. (I) is useful in gene therapy techniques to restore normal
XX activity of (II) or to treat disease states involving (II). (II) is
XX useful for generating antibodies against it, detecting or quantitating a
XX polypeptide in tissue, as molecular weight markers and as a food
XX supplement. (II) and its binding partners are useful in medical imaging
XX of sites expressing (II). (I) and (II) are useful for treating disorders
XX involving aberrant protein expression or biological activities. The
XX polypeptide and polynucleotide sequences have applications in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits to assess biodiversity
XX and to produce other types of data and products dependent on DNA and
XX amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic
XX amino acid sequences of the invention. Note: The sequence data for this
XX patent did not appear in the printed specification, but was obtained in
XX ftp.wipo.int/pub/published_pct_sequences
SQ Sequence 554 AA;
Query Match 8.7%; Score 249; DB 4; Length 554;
Best Local Similarity 21.6%; Pred. No. 3.7e-09;
Matches 125; Conservative 44; Mismatches 155; Indels 254; Gaps 27;
QY 10 PPSTS----LPQGSAPLP--SAPI-SSNMPQYTMQP-----QYVSPQHTLPPLQ 53
Db 11 PPAASRRVNLALGSCPLPLESSPAHRAVERPSVAHQPAVPGARRSLSEVPVLTVPALT 70
QY 54 PHHSQSPAPHSYMGQPPYRDLNRYPASSHDVYASSAAPIMPHTTVGSFLPPTSLSHNP 113
Db 71 P-----PRIPAMLTAVCGSL----- 85
QY 114 QAQAQAQSPHYPPPHSVLPPASSAQ-----SYQPIAPAP-----PRDRRADFNNG- 160

Db 86 --GSQTEAPHASPPRLDLQPLQTYQHTSPEAGDYPSLPQGLQSLPLGPEVDFSQY 143
QY 161 -----LPSGAFS-----YSDGKPOGW-----DPVAANGA--A 185
Db 144 ELPGASSRVTCEDLESPLAPGPFKLLQPDMSHHY-----ESWFRTHPGAEDGSWWD 198
QY 186 PYPG---XDSPRTQ-VVGSQRRG-----ILPSVPGR-- 213
Db 199 LHPGTWMDLPHTQGALTSPGHPGALQAGLGYYVDHQLCAPPPHHAHLLPAAGGQHL 258
QY 214 -----ATPVTNGVNGT-----GKNTTIPAKDADGKPPCPN----- 243
Db 259 LGPPDGALEVAAPESQGLDSSLDGAARPKGSRSSVPRSSGQTVCRCPNCLEAERLGAP 318
QY 244 -----CNKTYLHAKHLKRHLRHTGDRPYMC--VLCKDTFSRSDILK 283
Db 319 CGPDGKKKKHLHNCHIPGCGKAYAKTSHLKAHLRWHSGDRPFVCNWLFCGKRFRTRDELQ 378
QY 284 RHFQ-----KCSI-----RRGNPTGATHLSHP 305
Db 379 RHLQTHGTGKFPFCAVCSRVFMRSDHLAKMKTHEGAKEEAAGASGEGKAGGA--VEPP 436
QY 306 NAHVKRSQ--QAAANPVKPVQDEVSVSTVPPNGIP-----GTTYGEGAVNGNGLAP 355
Db 437 GGKGRKREAGSMASSPDSCDCFCVSVPPASAIPAVIFAHELGPTRGA--GGGVCP 494
QY 356 ARPGYADHQTMGFPMSSVNGMRGQPED----AFPGR 389
Db 495 RRHSHWEPAPGLXAVGGGGASGQAGDRGCHAXEAGR 532
RESULT 8
ADB64092
ID ADB64092 standard; protein; 723 AA.
XX
AC ADB64092;
XX 04-DEC-2003 (first entry)
DE Human protein encoded by clone BRAWH20176850.
XX Human; pharmaceutical; diagnostic; gene therapy; tissue regeneration;
KW cell regeneration; membrane protein; signal transduction-related protein;
KW transcription-related protein; osteoporosis; neurological disease;
KW cancer; tumour.
XX Homo sapiens.
XX EP1308459-A2.
XX 07-MAY-2003.
XX 28-MAR-2002; 2002EP-00007401.
XX 05-NOV-2001; 2001JP-00379298.
XX 25-JAN-2002; 2002US-00350978.
XX (HELI-) HELIX RES INST.
XX (REAS-) RES ASSOC BIOTECHNOLOGY.
XX Isogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S;
PI Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I;
PI Seki N, Yoshikawa T, Otsuka M, Nagahari K, Masuho Y;
XX WPI; 2003-450961/43.
XX N-PSDB; ADB62122.
XX New polynucleotides and polypeptides, useful for developing a diagnostic
PT marker or medicines for regulation of their expression and activity, or
PT as targets of gene therapy.
XX Claim 1; Page; 222pp; English.

CC The invention discloses a polynucleotide comprising a sequence selected
CC from 1970 fully defined nucleotide sequences which encode novel
CC polypeptides. Also claimed is a polypeptide encoded by the polynucleotide
CC or its partial peptide, an antibody binding to the polypeptide or peptide
CC of the polynucleotide, immunologically assaying the polypeptide or
CC peptide of the polynucleotide by contacting the polypeptide or peptide
CC with the antibody of the encoded protein, and observing the binding
CC between the two, a transformant carrying the polynucleotide in an
CC expressible manner and an antisense polynucleotide. The oligonucleotide
CC is useful as a primer for synthesising the polynucleotide, or as a probe
CC for detecting the polynucleotide. The polynucleotides and encoded
CC proteins are useful as pharmaceutical agents and many disease-related
CC genes may be included in them, for developing a diagnostic marker or
CC medicines for regulation of their expression and activity, or as targets
CC of gene therapy. The genes are involved in tissue and/or cell
CC regeneration. Membrane proteins, signal transduction-related proteins,
CC transcription-related proteins, disease-related proteins and genes
CC encoding them can be used as indicators for diseases (e.g. osteoporosis,
CC neurological diseases, cancer, tumours. The cDNA may be used to regulate
CC the activity or expression of the encoded protein to treat diseases. The
CC sequence presented is a protein of the invention. Note: Some of the
CC sequence data for this patent is not represented in the printed
CC specification, but is based on sequence information supplied by the
CC European Patent Office.

XX Sequence 723 AA;

Query Match 8.4%; Score 240; DB 7; Length 723;
Best Local Similarity 22.6%; Pred. No. 2.1e-08;
Matches 132; Conservative 54; Mismatches 198; Indels 200; Gaps 27;

QY 19 SAPLSPAPISMPMPQYTMQPPVPSQPHLPPLQPHHSQSPAPHSYMGQPPYPDLNRY 78
Db ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
84 SMPPPFVPSQMPPPPLPTMPP--PVLPPSLPPPPVMP--PALPATVPPPGMPP----- 131
QY 79 PASSHDVVASSAAPIMPHTTVGSLLPPTSLSHPNPQAAQAAQSQSPHYPPPH-----SV 131
Db :
132 -----PVMPPSLPTSVPPPGM-----PPSLSSAGPPPVLPPLPSLFSAGPPPV 173
QY 132 LPPASSAQSPQPIAPAPPRDRRADFNGLPSGAFSYSDGKPGQWDPVAANGAAPYPGKD 191
Db ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
174 LPPPSLSSTAPPVMPPLPLLS-SATPPPGIP-----PPGVPQGIPLQLT--AAPVPPAS 224
QY 192 SPRTQVVGSGRRGILPSVPGRATPVTVNGVTGKNTTI-----PAKDADGKFPCPN 243
Db :
225 SSQSSQVPEKRPALLP-----TPVS---FGSAPPTTYHPPLQSGAFSEQVNSKAP--- 272
QY 244 CNKTYLHAKHLKRHLRHTGDRPYMCVLCCKDTFSRSDILKRHFQKCSIRRGNPTGATHLS 303
Db : : : : : ||||| :
273 -----LSKSAL-----PY-----SSFSSDQGL----- 289
QY 304 HPNAHVKRSQQAAANPVKPVQD---EVSSTVPPPNGIPGTTYGEG----- 346
Db :
290 -----GESSAAPSQPITAVKDMFVRSGLLPDP---PRSSYLESPRGPRFDGPRRFED 339
QY 347 -----AVNGNGLAPARPGYADHQTMG-----FPMSSVNG 375
Db ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
340 LGSRCGPRPKGRPFEGNRPDGPRPRYEGHPAEGTCKWGMIPRGPASQFYITPSTSLSP 399
QY 376 MGRGQ---PEDAPGGRPHQGAPWPQAPKQSPYLVPQGADPSGHQ-----LNID 421
Db | : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : : : : : : : :
400 RQSGPQWKGPAPAF--GQHQHQKQPKSQAEPLSGN-KEPLADTSSNQKNFKMQSAAFSIA 456
QY 422 RNIEQVK-----QPVVQDPKRPVMPGHPGHPGELDW-TSMFQPAPEGYMFWSQSMFG 472
Db : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : : : : : : : :
457 ADVKDVKAAQSNENLSDSQEQEPKSEVSEGPVPSNWDQNVQSMETQIDKAQAVTQPVPL 516
QY 473 GQEPHIAHVETERKYPTTTAGQESGMNGLYLASTMSGDGTVPQ 516
Db :
517 ANKPVPA-----QSTFPSTKGGMEGGT--AVATSSLTADNDFKP 553

ADC31187
ID ADC31187 standard; protein; 2062 AA.
XX
AC ADC31187;
XX
DT 18-DEC-2003 (first entry)
XX
DE Human novel polypeptide sequence, SEQ ID NO:1269.
XX
KW Human; diagnostic; drug screening; forensics; gene mapping;
KW biodiversity assessment; Parkinson's disease; Alzheimer's disease;
KW neurodegenerative diseases; anaemia; platelet disorder; wound; burns;
KW ulcers; osteoporosis; autoimmune disease; cancer;
KW molecular weight marker; food supplement; antiparkinsonian; nootropic;
KW neuroprotective; antianaemic; anticoagulant; thrombolytic; vulnerary;
KW antiulcer; osteopathic; immunosuppressive; antiinflammatory; cytostatic;
KW gene therapy; chromosome 14q24.3.
XX
OS Homo sapiens.
XX
PW WO2003029271-A2.
XX
PD 10-APR-2003.
XX
PF 24-SEP-2002; 2002WO-US030474.
XX
PR 24-SEP-2001; 2001US-0324631P.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang TY, Zhang J, Ren F, Xue AJ, Zhao QA, Wang J, Wehrman T;
PI Zhou P, Ghosh M, Wang D, Ma Y, Asundi V, Wang Z, Weng G;
PI Haley-Vicente D, Drmanac RT;
XX
DR WPI; 2003-371981/35.
DR N-PSDB; ADC30216.
XX
PT New polynucleotide and polypeptide useful for diagnosing, preventing or
PT treating conditions such as neurodegenerative diseases, anemias, platelet
PT disorders, wounds, burns, ulcers, osteoporosis, autoimmune diseases or
PT cancer.
XX
PS Claim 20; SEQ ID NO 1269; 1185pp; English.
XX
CC The invention relates to 971 novel human cDNA sequences (ADC29919-
CC ADC30889) and the polypeptides they encode (ADC30890-ADC31860). The
CC invention also relates to nucleic acid sequences over 99% identical with
CC the novel human cDNAs. The invention additionally encompasses expression
CC vectors and host cells comprising a nucleic acid of the invention; the
CC recombinant production of a polypeptide of the invention; an antibody
CC against a polypeptide of the invention; a method of detecting
CC polynucleotides or polypeptides of the invention; and methods of
CC identifying a compound which binds to a polypeptide of the invention. The
CC invention further discloses methods of preventing, treating or
CC ameliorating a medical condition; kits comprising polynucleotide probes
CC and/or monoclonal antibodies for carrying out the methods of the
CC invention; methods for the identification of compounds that modulate the
CC expression or activity of the polynucleotide and/or polypeptide; and 767
CC contig sequences corresponding to the cDNA sequences of the invention
CC (ADC31861-ADC32627) and the polypeptides encoded by the contigs (ADC32628
CC -ADC33394). The nucleic acids and polypeptides of the invention are
CC useful in diagnostics, drug screening, forensics, gene mapping, in the
CC identification of mutations responsible for genetic disorders or other
CC traits, for assessing biodiversity, and in producing many other types of
CC data and products dependent on DNA and amino acid sequences. They are
CC also used for treating diseases such as Parkinson's disease, Alzheimer's
CC disease and other neurodegenerative diseases, anaemia, platelet
CC disorders, wounds, burns, ulcers, osteoporosis, autoimmune diseases or
CC cancer. The nucleic acids may also be used as hybridisation probes or
CC primers, and in the recombinant production of a protein. The polypeptides
CC are also useful in generating antibodies, as molecular weight markers,
CC and as food supplements. The present sequence represents a specifically
CC claimed human polypeptide sequence of the invention. Note: The sequence

CC data for this patent did not form part of the printed specification, but
CC was obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published pct sequences.

Db 1093 HHGSAP--PPTSAHQSPWFCSTSYICSSITTVLLHLLHLLHHGSAPP-----PTSAHHH 1146

QY 393 G-APWPQAPKQSP 404

Db 1147 GSAPPPTSAHQSP 1159

RESULT 11

ADC07968

ID ADC07968 standard; protein; 1142 AA.

XX

AC ADC07968;

XX

DT 18-DEC-2003 (first entry)

XX

DE Rice protein sequence Seq ID234 related to grain filling.

XX plant biotechnology; carbohydrate synthesis; carbohydrate metabolism;

KW carbohydrate degradation; carbohydrate; plant grain; grain filling; corn;

KW tomato; banana; canola; cotton; peanut; sorghum; tobacco; sugarbeet;

KW wheat; rice; protein; oil; starch; fibre; moisture content; cereal grain;

KW gene; ds; plant.

XX Oryza sativa.

OS

XX WO2003000905-A2.

PN

XX 03-JAN-2003.

PD

XX 21-JUN-2002; 2002WO-IB002450.

PF

XX 22-JUN-2001; 2001US-0300112P.

PR

PR 26-SEP-2001; 2001US-0325277P.

PR

PR 20-DEC-2001; 2001US-0342327P.

XX

XX (SYGN) SYNGENTA PARTICIPATIONS AG.

PA

XX Zhu T, Cheng W, Briggs S, Cooper B, Goff SA, Moughamer T;

PI Glazebrook J, Katagiri F, Kreps J, Provart N, Ricke D;

PI

XX WPI; 2003-229341/22.

DR

DR N-PSDB; ADC07967.

XX

XX New plant genes encoding polypeptides having an activity involved in or

PT associated with the synthesis, metabolism or degradation of carbohydrates

PT in the plant grain useful in generating plants having improved

PT nutritional properties.

XX Claim 15; SEQ ID NO 234; 130pp; English.

PS

XX This invention, in the area of plant biotechnology, relates to novel

CC polynucleotides comprising a nucleotide sequence encoding a protein which

CC is involved in or associated with the synthesis, metabolism or

CC degradation of carbohydrates in the plant grain and the expression of

CC which is up-regulated during grain filling. The plant is selected from

CC corn, tomato, banana, canola, cotton, peanut, sorghum, tobacco,

CC sugarbeet, wheat, and rice. The invention may be useful for the

CC improvement of protein, oil, starch, fibre and moisture content of the

CC cereal grains. In addition, carbohydrate levels may be modified to a more

CC desirable level using the present invention. The present sequence is the

CC amino acid sequence of a rice protein of the invention. Note: The

CC sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from WIPO

CC at ftp.wipo.int/pub/publishedpct_sequences.

XX Sequence 1142 AA;

SQ

Query Match 8.3%; Score 238.5; DB 7; Length 1142;

Best Local Similarity 23.2%; Pred. No. 4.6e-08;

Matches 145; Conservative 56; Mismatches 243; Indels 181; Gaps 30;

QY 3 PRNHP-----SRPPSTSLPQGSAPLPSAPISSMPMPQYTMQPQYPVSPHTLPPLQPHHS 57

Db 248 PQNQPHLTQLCAPAAQSQSQHPPPLQSAQAQTQIVQAQPSQVSFQQPQPHAQPT-----T 302

QY 58 QSPAPHSYMGOPPYRDLNRYPASSHDVYASSAAPIMPHTTVGSLLPFTSLSHNPQAA 117

Db 303 HTPVPTQLGSQPFAMPPTQATPSEHHQVMAQQQPQLQH-----LPOQ---QHENAQQQS 354

QY 118 QAQ-QSPHYPPPHSVLPASSA-----QSYPOPIAPAPPRDRRADFNGL-----P 162

Db 355 YPQMQAQYHQPPMSHAQPNPSVHAVTGHQSFSQP-QPAHQMQQGAQLQSLHVSQQQMP 413

QY 163 SGAFSYSDGKPGQWDP-VAANGAAPYPGKDSPTQVWGSQG-RRGILPSVPGRATPVING 220

Db 414 SAQHHAHAHTPQGQOPTMAQGI-----QQTPOHQHVGHHLRPEIVASIPPAAP----- 464

QY 221 VNGTGKNTTIPAKDADGKFFPCPNCKNTYLHA-----KHLKRHLRHHTGDRPYMCVLCK 273

Db 465 -QGFFLNAPAPSQTGQSYQQGMPSSQQLMHAPLQSQGQFMQQHPTHTSAGRSMNYVAPQ 523

QY 274 DTFSR-----SDILKRHFQKC- 289

Db 524 EQFQNSGGPVKGLQAGVMNQPPMRMASDNVGTSELHGAGQSFQGGSSSLKKPTSESE 583

QY 290 -SIRRGNPTGATHLSHPNAHVKRSSQQAANPV-----KPVQD-----EVSST 331

Db 584 KSENATNGTNGTEVSGKNG---SAESALVNPISLDGSDGSKGKGKGVDFSAWESNSH 639

QY 332 VPPPNGIGTTYG--EGAVNGNL--APARPGYAD-----HQTMGFFMSSVNGMGRGQ-- 380

Db 640 DPDARGGKGRSGISNDLVKGSLLQQAPOHPYGPDSMLPQHMRQPGHMPYMQGLPNQMRP 699

QY 381 PEDAFP-GGRPHQGAPWPQAPKQSPYLVPQGADPSGHQNLNIDRNIEQVKQPVVQDPKRPV 439

Db 700 PKHSFEPNSRPPMQPFEMAPR-----VPG--PNQNMQISQSI-RPDGAIVRPPMGAP 750

QY 440 MPG-HPGHGELDWTSMFQPAPE-----GYMFSQ----- 468

Db 751 MPGLH-----DSTVPPFAPEYVGTKKNNSVGNPGHGSRALFEGGFNSSQKHKSC 801

QY 469 -SMPGGQEPITHAHVETERKYYP TTT 492

Db 802 AANPGRNNVSHKDFEDNMKQFPVPT 826

RESULT 12

ABB65293

ID ABB65293 standard; protein; 1192 AA.

XX

AC ABB65293;

XX

DT 26-MAR-2002 (first entry)

XX

DE Drosophila melanogaster polypeptide SEQ ID NO 22671.

XX

KW Drosophila; developmental biology; cell signalling; insecticide;

KW pharmaceutical.

XX

OS Drosophila melanogaster.

XX

PN WO200171042-A2.

XX

PD 27-SEP-2001.

XX

PF 23-MAR-2001; 2001WO-US009231.

XX

PR 23-MAR-2000; 2000US-0191637P.

PR 11-JUL-2000; 2000US-00614150.

XX

PA (PEKE) PE CORP NY.

XX

PI Venter JC, Adams M, Li PWD, Myers EW;

XX

DR WPI; 2001-656860/75.

Db 425 AEAKEQAAAAIKLEKKEKSGKPLTPPVFEKQEQPDTPLVNYAPYANLYQHSTAGSSV 484

QY 354 APARP 358

Db 485 NPPPP 489

RESULT 14

AAAY33497

ID AAY33497 standard; protein; 1185 AA.

XX

AC AAY33497;

XX

DT 19-JAN-2000 (first entry)

XX

DE Human atrophin I protein.

XX

KW Proapoptotic; dependence domain; p75NTR; androgen receptor; DCC;

KW huntingtin polypeptide; Machado-Joseph disease; SCA1; SCA2; SCA6;

KW atrophin-1; cell death; apoptosis; Huntington's disease; head trauma;

KW Alzheimer's disease; Kennedy's disease; spinocerebellar ataxia; stroke;

KW dentatorubropallidolysian atrophy; cell proliferation; cell survival;

KW neoplastic; malignant; autoimmune; fibrotic.

XX

OS Homo sapiens.

XX

PN WO9945944-A1.

XX

PD 16-SEP-1999.

XX

PF 11-MAR-1999; 99WO-US005250.

XX

PR 12-MAR-1998; 98US-00041886.

XX

PA (BURN-) BURNHAM INST.

XX

PI Bredesen DE, Rabizadeh S;

XX

DR WPI; 1999-561617/47.

DR N-PSDB; AAZ23430.

XX

PT New proapoptotic dependence peptides, used to develop products for

PT treating, e.g. Alzheimer's disease.

XX

PS Disclosure; Page 154-157; 199pp; English.

XX

CC This invention describes novel pure proapoptotic dependence peptides

CC which comprise a sequence of an active dependence domain selected from

CC dependence polypeptides consisting of p75NTR, androgen receptor, DCC,

CC huntingtin polypeptide, Machado-Joseph disease gene product, SCA1, SCA2,

CC SCA6 and atrophin-1 polypeptide. The proapoptotic peptides are capable of

CC inducing cell death and can be used to develop products to mediate or

CC inhibit apoptosis. The methods can be used for reducing the severity of a

CC proapoptotic dependence domain mediated pathological conditions e.g.

CC Huntington's disease, Alzheimer's disease, Kennedy's disease,

CC Spinocerebellar ataxias, dentatorubropallidolysian atrophy, Machado-

CC Joseph disease, stroke or head trauma. They can also be used for reducing

CC the severity of a pathological condition mediated by upregulated cell

CC proliferation or cell survival e.g. neoplastic, malignant, autoimmune or

CC fibrotic conditions. This sequence represents the human atrophin I

CC polypeptide described in the method of the invention

XX

SQ Sequence 1185 AA;

Query Match 8.1%; Score 233.5; DB 2; Length 1185;

Best Local Similarity 22.0%; Pred. No. 1.1e-07;

Matches 123; Conservative 46; Mismatches 193; Indels 197; Gaps 23;

QY 7 PSRPPSTSLPQGSAPLPSPAPISSMPMPQYTMQPQYPVVSQPHTLPLPQPHHSQSPAPHSYM 66

Db 272 PTKPPTT--PVGGNLPSAP-----PPANFPHVTPNLPP----- 303

QY 67 GQPPYRPDLNRYPASSHDVYASSAAPIMPHITVGSLPPTSFLSHRNPPQAQA----- 119

Db 304 --PPALRPLNNASASPPGL---GAQPLPGHL-----PSPYAMQGGMGLPPG 345

QY 120 -QQSPHY-PPPHSVLPASSAQSYPOPIAPAPRRDRADFNGLPSGAFSYSGKPKQWD 177

Db 346 PEKGPTLAPSHS-LPPASSS-----APAPMRFPYSSSSSSSSSSSSSSSSAS 397

QY 178 PVAANGAAP-YPGKDSPTQV-VGSQGRRGILPSVPGRATPVTNGVGTGKNTTIPAKDA 235

Db 398 PFPASQALPSYPHSFPPTSLSVSNQPKYQTQPSLPSQAV-----WS 439

QY 236 DGKFFPCPNCKTYLHAKHLKRLLRHTGDRPYMCVLCCKDTFSRSDILKRHFQKCSIRRG 295

Db 440 QGPPPPPPYGRLLANSN-----AHPGFPF-----P 464

QY 296 PTGATHLSHP---NAHVKRSQQQAAANPVQDEVSVSTVPPNGIPGTTYGEGAVNGN- 351

Db 465 STGAQSTAHPPVSTHHHHHQQQQQQQQQQQQHHGNSGPPPPGAFPHPLEGGSSHHAHP 524

QY 352 -----GLAPARPG-----YADHQTMTGFFPMSSVNGMGRGQPEDAFPPGGRPH 391

Db 525 YAMSPSLGSLRYPYPGPAHLPPPHSQVSYSQAGPNGPPVSSSSSSSTSQSYPCSHPS 584

QY 392 -----QGAPWQAP-----KQSPYLVQPGADPSGHQLNIDRNIEQVKQPVVQDPKR 437

Db 585 PSQGPQGAPYPPVPVTVTTSATLSTVIATVASSPAGY-----KT 625

QY 438 PVMPGHPGHPGELDWTSMFQPAPEGYMFSQSMPPGGQEPHHAHVETERKYVPTTAGQES 497

Db 626 ASPPGPPPYCKRAPSPGAYKTATPPGY-----KPGSP-----PSFRTGTTPP 666

QY 498 GMNGLYLASTMSGDGTVPQ 516

Db 667 GYRG---TSPFAGPGTFKP 682

RESULT 15

ABB65135

ID ABB65135 standard; protein; 926 AA.

XX

AC ABB65135;

XX

DT 26-MAR-2002 (first entry)

XX

DE Drosophila melanogaster polypeptide SEQ ID NO 22197.

XX

KW Drosophila; developmental biology; cell signalling; insecticide;

KW pharmaceutical.

XX

OS Drosophila melanogaster.

XX

PN WO200171042-A2.

XX

PD 27-SEP-2001.

XX

PF 23-MAR-2001; 2001WO-US009231.

XX

PR 23-MAR-2000; 2000US-0191637P.

PR 11-JUL-2000; 2000US-00614150.

XX

PA (PEKE) PE CORP NY.

XX

PI Venter JC, Adams M, Li PWD, Myers EW;

XX

DR WPI; 2001-656860/75.

DR N-PSDB; ABL09238.

XX

PT New isolated nucleic acid detection reagent for detecting 1000 or more

PT genes from Drosophila and for elucidating cell signaling and cell-cell

PT interactions.

XX

PS Disclosure; SEQ ID NO 22197; 21pp + Sequence Listing; English.

CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
CC ABB72072). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 926 AA;

Query Match		8.0%;	Score 229.5;	DB 4;	Length 926;
Best Local Similarity		22.9%;	Pred. No. 1.5e-07;		
Matches 142;		Conservative 44;	Mismatches 229;	Indels 205;	Gaps 30;
QY	4 RNHP-----SRPPSTSLP-QGSAPLPS-----APISSMEMPPQYTMQP	39			
Db	351 QQHPVYRNAQGGQPGAGQVFGQGGPVQSVINPNAAPNQRPNNGLSGPNPQQQQQQ	410			
QY	40 QYPVSPHTLPLQPHHSQSPAPHSYMGQPPYRDLNRYPASSHDVYASSAAPIMPHTTV	99			
Db	411 PPGGQQ--PPNQQQQQQTGP---GGPGQP-----GAGGPGVP----	445			
QY	100 GSLPPTSLSHPNPQAQAQAQSQSPHYPP-----PHSVLPASSAQSYPOPIAPAPPRD	152			
Db	446 ---PPQSPYRVSYQQQQQSHYPGYPPQPQTQYQPGAYPYGPPTQGYGPP-PPGPPNA	501			
QY	153 RRADFNGLPSGAFSYSDGKPKQGWDFVAANGAARYPGKDSPTQVVSGRRGILPSVPG	212			
Db	502 AQGGYHGG-PAGATGASG--HGYPNAGAGQGGPPPGAYPPPP---GSQQ-----VPPVPG	551			
QY	213 RATPVTN----GVNGTGKNTTIPAKDADGKF-PCPNCNKTYLHAKHLKRHLRHTGDRPY	267			
Db	552 QQQPPPGPPPPGPPPTGGQQQPPPPGPPPSQYGGPPPPQNS-----AGGPPP	596			
QY	268 MCVLCKDTFSRSDILKRHFQKCSIRRGNTG-----ATHLSHPNAHVKRSQQQAAA	318			
Db	597 M-----GYAGYPNPGYQYQAGAGGGPPPSGYWPPPPPTSSAQSPYQAYQQQQQAAA	650			
QY	319 N-----PVKPVQDEVST-----VPPPNGIPGTYGEGAVN----	349			
Db	651 GGGAGAPPSSYPGGPPTSGAAPPPPPGGAYSTTAPSTQTPPPQGGGAGGGNNPNPNA	710			
QY	350 -----GNGLAPARPGYADHQTMGFMSSVNGMGRGQPEDAFPGGRPHQG-APWP	397			
Db	711 QQSTPPPGGAGGAGSGPGGAGQQYAGPPQ-----QPQQQPPGVVSVGVAPLP	762			
QY	398 ---QAPKQSP--YLVQPGADPSGHQLNIDRNIEQVKQPVW-----	432			
Db	763 TQVQPTYSTPGNYNQPGAPPNNQQQQQQQQQQQQQQTTPPSAGGSAGGGGAPNAQGGQ	822			
QY	433 -QDP----KRPVMPGHPGHPGELDWTSMFQQAPEG-----YMFQSMPGGQEPIHA	479			
Db	823 NQQPPPNGATPPMPNQQYQPA-----PGAPQGPYGGPPPPQAYGPPPPGSAYPGHA	873			
QY	480 HVETER----KYYPPTTAGQ	495			
Db	874 YHQPQAGGYAQYPTQGYQ	893			

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 1, 2004, 17:39:33 ; Search time 17 Seconds
(without alignments)
1589.671 Million cell updates/sec

Title: US-10-029-180-8
Perfect score: 2866
Sequence: 1 MDPNRNHPSPSTSLPQGSA.....NGLYLASTMSGDGTGVPARQ 519

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0
Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	255.5	8.9	722	1	Z219_HUMAN	Q9p2y4 homo sapien
2	251.5	8.8	2716	1	OSA_DROME	Q8in94 drosophila
3	236.5	8.3	644	1	BTD_DROME	Q24266 drosophila
4	234.5	8.2	1185	1	DRPL_HUMAN	P54259 homo sapien
5	233	8.1	856	1	PRD1_MOUSE	Q60636 mus musculus
6	225	7.9	692	1	STEA_EMENI	O74252 emericeella
7	223	7.8	470	1	EGR2_MOUSE	P08152 mus musculus
8	222	7.7	963	1	YQ36_CAEEL	Q09457 caenorhabdi
9	221.5	7.7	431	1	SP7_HUMAN	Q8tddd2 homo sapien
10	221.5	7.7	582	1	YANB_SCHPO	Q10076 schizosacch
11	221	7.7	470	1	EGR2_RAT	P51774 rattus norv
12	218.5	7.6	5262	1	MLL2_HUMAN	O14686 homo sapien
13	212.5	7.4	1183	1	DRPL_RAT	P54258 rattus norv
14	211.5	7.4	354	1	KLF2_MOUSE	Q60843 mus musculus
15	211	7.4	789	1	PRD1_HUMAN	O75626 homo sapien
16	210.5	7.3	543	1	EGR1_HUMAN	P18146 homo sapien
17	209.5	7.3	533	1	EGR1_MOUSE	P08046 mus musculus
18	209	7.3	398	1	SP5_MOUSE	Q9jbx2 mus musculus
19	208.5	7.3	323	1	KLFE_HUMAN	Q8td94 homo sapien
20	207.5	7.2	1386	1	ZAP3_MOUSE	Q9r0i7 mus musculus
21	207	7.2	476	1	EGR2_HUMAN	P11161 homo sapien
22	206.5	7.2	508	1	EGR1_RAT	P08154 rattus norv
23	206.5	7.2	5120	1	PCLO_CHICK	Q9pu36 gallus gall
24	206	7.2	428	1	SP7_MOUSE	Q8vi67 mus musculus
25	206	7.2	1157	1	SRA4_HUMAN	O95104 homo sapien
26	205.5	7.2	345	1	KLF3_HUMAN	P57682 homo sapien
27	205.5	7.2	5147	1	PCLO_HUMAN	Q9y6v0 homo sapien
28	202	7.0	1822	1	ZAP3_HUMAN	P49750 homo sapien
29	201.5	7.0	1300	1	SAL3_HUMAN	Q9bxa9 homo sapien
30	201.5	7.0	5085	1	PCLO_RAT	Q9jks6 rattus norv
31	200.5	7.0	446	1	KLF5_MOUSE	Q9z0z7 mus musculus
32	200	7.0	362	1	KLF1_HUMAN	Q13351 homo sapien
33	199.5	7.0	481	1	Z358_HUMAN	Q9nw07 homo sapien

34	199.5	7.0	511	1	EGR1_BRARE	P26632 brachydanio
35	198.5	6.9	449	1	WT1_PIG	O62651 sus scrofa
36	198.5	6.9	4903	1	MLL3_MOUSE	Q8brh4 mus musculus
37	198	6.9	344	1	KLF3_MOUSE	Q60980 mus musculus
38	198	6.9	251	1	KLF2_RAT	Q9et58 rattus norv
39	197.5	6.9	421	1	EGR2_XENLA	Q08427 xenopus lae
40	197.5	6.9	780	1	YNL5_CAEEL	Q21955 caenorhabdi
41	197	6.9	432	1	BRLA_EMENI	P10069 emericeella
42	196.5	6.9	477	1	MAZ_HUMAN	P56270 homo sapien
43	195.5	6.8	449	1	WT1_HUMAN	P19544 homo sapien
44	194.5	6.8	467	1	CBPA_DICDI	P35085 dictyosteli
45	194.5	6.8	773	1	Z341_HUMAN	Q9byn7 homo sapien

ALIGNMENTS

RESULT 1
Z219_HUMAN
ID Z219_HUMAN STANDARD; PRT; 722 AA.
AC Q9P2Y4; Q9BW28;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Zinc finger protein 219.
GN ZNF219.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., AND CHARACTERIZATION.
RC TISSUE=Testis;
RX MEDLINE=20277481; PubMed=10819330;
RA Sakai T., Toyoda A., Hashimoto K., Maeda H.;
RT "Isolation and characterization of a novel zinc finger gene, ZNF219,
and mapping to the human chromosome 14q11 region.";
RL DNA Res. 7:137-141(2000).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Kidney;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
Klausner R.D., Collins F.S., Wagner L., Schenmen C.M., Schuler G.D.,
Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,
Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
Bosak S.A., MCEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
Villalon D.K., Muzny D.M., Sodergren E.J., Iu X., Gibbs R.A.,
Fahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC -!- FUNCTION: May function as a transcription factor.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- TISSUE SPECIFICITY: Ubiquitous.
CC -!- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-FINGER PROTEINS.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----
DR EMBL; AB015427; BAA90526.1; --
DR EMBL; BC000694; AAH00694.1; --
DR Genew; HGNC:13011; ZNF219.
DR MIM; 605036; --
DR GO; GO:0005634; C:nucleus; TAS.
DR GO; GO:0003700; F:transcription factor activity; TAS.
DR GO; GO:0006351; P:transcription, DNA-dependent; TAS.
DR InterPro; IPR007087; Znf_C2H2.
DR Pfam; PF00096; zf-C2H2; 8.
DR SMART; SM00355; Znf_C2H2; 9.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 5.
DR PROSITE; PS0157; ZINC_FINGER_C2H2_2; 6.
KW Transcription regulation; DNA-binding; zinc-finger; Metal-binding;
KW Nuclear protein; Repeat.
FT ZN_FING 57 79 C2H2-TYPE.
FT ZN_FING 85 107 C2H2-TYPE.
FT ZN_FING 163 186 C2H2-TYPE.
FT ZN_FING 274 296 C2H2-TYPE.
FT ZN_FING 302 324 C2H2-TYPE.
FT ZN_FING 498 520 C2H2-TYPE.
FT CONFLICT 232 233 MISSING (IN REF. 2).
FT CONFLICT 436 436 Q -> E (IN REF. 2).
SQ SEQUENCE 722 AA; 76876 MW; B19DA77B148BC45B CRC64;

Query Match 8.9%; Score 255.5; DB 1; Length 722;
Best Local Similarity 24.3%; Pred. No. 3.7e-06;
Matches 145; Conservative 46; Mismatches 156; Indels 249; Gaps 35;
QY 7 PSRPPSTSLPQGSAPLPSAPISSMPMPQYTMQPQYFVSQPHTLPPLOPHHSQSPAP-HSY 65
Db 215 PERPLAAT---SAAPPP-----QPQPPPPQPE-PRSVQPEPEPEPEREATPTPAPAA 264
QY 66 MGQPPYRPD-----LNRYPAS-----SH-DVY 86
Db 265 PEEPPAPPEFCQVCGSQFTQSWFLKGHMRKHKASFHDACPVCGRCFKEPFWFLKNHMKVH 324
QY 87 ASSAAPIMPHTVGSLLPPTSFLSHPNP-QAQAQAQSQSPH-----YPPPHSVLPASSAQS 140
Db 325 ASKLG-----LRAPGASGARAPQPPDLGLLAYEP-----LGPALL--- 362
QY 141 YPQIAPAP-PRDRADFNGLPSGAFSY-----SDGKPQGWDPVAANGAAPPYPGKD--- 191
Db 363 ----LAPAPTAEERRE-----PPSLGLVLSLRAGEGRPNG-----EGAEPGGRSFGG 406
QY 192 -SPRTQVVGSGQRR-----GILPSV-----PGR----- 213
Db 407 FRPLSSALPARARRHRAEPEEEEEVEVQAQETWARGSLGSLASLHPRPCEGPGHSASA 466
QY 214 -----ATPVTVNG--VNGT---GKNTTIPAKDADGKFCPCNCKTYLHAKHLKRHL 259
Db 467 AGAQAARSTATQEENGLLVGTRPEGGRGAT--GKD-----CPFCGKSFRSAHHLKVHLR 518
QY 260 RHTGDRPYMCVLCKDTFSRSDILKRHFQKCSIRGNPTGATHLSHFNHVKRS---QQQA 316
Db 519 VHTGERPYKCPCHDYAGTQSGSLKYHLQRHHREQRSAGPGPPPEPPPSQSGAPQSGA 578
QY 317 AANPKFVQDEVSSTVPPPNIGIPGTTYGEGAVNGNGLAPARPGYADHOTMGFFPMSSVNGM 376
Db 579 KPSPQPATWVEGASSPRPPS-----SGAGPGSRRKPPASPG---RTLRL-----NGR 620
QY 377 GRGQPE-----DAFPGGRPHQ-----GAPW-----P 397
Db 621 G-GEAEPLDLSLRAGPGGEAGPGGALHRCFLFCPATGAPELMALHLQVHHSRRARRRRPP 679
QY 398 QAPKQSPYLVPQGDPSGHQLNIDRNIEQVKQPVVQDPKRPVMPGHPG---HPGE 449
Db 680 QADASFPY---ARVPSG-----ETPPSPSQEGEGSGLSRPGE 714

OSA_DROME STANDARD; PRT; 2716 AA.
ID OSA_DROME
AC Q8IN94; O61603; Q9VEG7;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Trithorax group protein OSA (Eyelid protein).
GN OSA OR ELD OR CG7467.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A., FUNCTION, SUBCELLULAR LOCATION, TISSUE
RX SPECIFICITY, AND DEVELOPMENTAL STAGE.
MEDLINE=97415319; PubMed=9271118;
RA Treisman J.E., Luk A., Rubin G.M., Heberlein U.;
RT "eyelid antagonizes wingless signaling during Drosophila development
and has homology to the Bright family of DNA-binding proteins.";
RL Genes Dev. 11:1949-1962(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkeley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazew R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Douc L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodex A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
RN [3]
RP FUNCTION, AND DEVELOPMENTAL STAGE.
RX MEDLINE=99112962; PubMed=9895321;
RA Vazquez M., Moore L., Kennison J.A.;
RT "The trithorax group gene osa encodes an ARID-domain protein that
genetically interacts with the brahma chromatin-remodeling factor to
regulate transcription.";
RL Development 126:733-742(1999).
RN [4]
RP DNA-BINDING, AND IDENTIFICATION IN A BRAHMA COMPLEX WITH BRM AND SNR1.

RX MEDLINE=20069333; PubMed=10601025;
RA Collins R.T., Furukawa T., Tanese N., Treisman J.E.;
RT "Osa associates with the Brahma chromatin remodeling complex and
RT promotes the activation of some target genes.";
RL EMBO J. 18:7029-7040(1999).
RN [5]
RP FUNCTION.
RX MEDLINE=99403006; PubMed=10471712;
RA Staehling-Hampton K., Ciampa P.J., Brook A., Dyson N.;
RT "A genetic screen for modifiers of E2F in Drosophila melanogaster.";
RL Genetics 153:275-287(1999).
RN [6]
RP IDENTIFICATION IN A BRAHMA COMPLEX WITH BRM; OSA; MOR; SNR1; DALAO;
RP BAP55; BAP60 AND BAP47, AND FUNCTION AS COACTIVATOR.
RX MEDLINE=20270023; PubMed=10809665;
RA Kal A.J., Mahmoudi T., Zak N.B., Verrijzer C.P.;
RT "The Drosophila brahma complex is an essential coactivator for the
RT trithorax group protein zeste.";
RL Genes Dev. 14:1058-1071(2000).
RN [7]
RP FUNCTION AS A COREPRESSOR.
RX MEDLINE=20573925; PubMed=11124806;
RA Collins R.T., Treisman J.E.;
RT "Osa-containing Brahma chromatin remodeling complexes are required
RT for the repression of wingless target genes.";
RL Genes Dev. 14:3140-3152(2000).
RN [8]
RP FUNCTION AS A COREPRESSOR, AND INTERACTION WITH PNR AND CHI.
RX MEDLINE=22515897; PubMed=12629041;
RA Heitzler P., Vanolst L., Biryukova I., Ramain P.;
RT "Enhancer-promoter communication mediated by Chip during
RT Pannier-driven proneural patterning is regulated by Osa.";
RL Genes Dev. 17:591-596(2003).
CC -!- FUNCTION: Trithorax group (trxG) protein required for embryonic
CC segmentation, development of the notum and wing margin, and
CC photoreceptor differentiation. Required for the activation of
CC genes such as Antp, Ubx and Eve. Binds to DNA without specific
CC affinity, suggesting that it is recruited to promoters by
CC promoter-specific proteins. Essential component of the Brahma
CC complex, a multiprotein complex which is the equivalent of the
CC yeast SWI/SNF complex and acts by remodelling the chromatin by
CC catalyzing an ATP-dependent alteration in the structure of
CC nucleosomal DNA. This complex can both serve as a transcriptional
CC coactivator or corepressor, depending on the context. Acts as an
CC essential coactivator for Zeste, which recruits the whole complex
CC to specific genes. In contrast, it acts as a corepressor for Wg
CC target genes, possibly via an interaction with Pan and Gro. It
CC also acts as a negative regulator for proneural achaete-scute,
CC when it is directly recruited by Pan and Chi. Also represses E2f
CC activation.
CC -!- SUBUNIT: Component of the Brahma complex, which is composed of
CC Brw, Osa, Mor, Snr1/Bap45, Bap11/Dalao, Bap55, Bap60 and Bap47.
CC Interacts with Pnr and Chi via its EHD domain.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- TISSUE SPECIFICITY: Ubiquitously expressed in early embryo. In
CC third instar larvae, it is ubiquitously expressed in wing and eye-
CC antenna imaginal disks, with a stronger expression in a band just
CC anterior to the morphogenetic furrow.
CC -!- DEVELOPMENTAL STAGE: Expressed both maternally and zygotically.
CC -!- DOMAIN: The ARID domains mediates the binding to DNA.
CC -!- SIMILARITY: Contains 1 ARID domain.
CC -!- SIMILARITY: Contains 1 EHD (Eyelid homology) domain.
CC -!- CAUTION: Ref.2 (AAF55457) sequence differs from that shown due to
CC erroneous gene model prediction.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; AF053091; AAC06254.1; -.
DR EMBL; AE003718; AAF55457.1; ALT_SEQ.
DR EMBL; AE003718; AAN13750.1; -.
DR PIR; T13049; T13049.
DR FlyBase; FBgn0003013; osa.
DR GO; GO:0005634; C:nucleus; IDA.
DR GO; GO:0003677; F:DNA binding; IDA.
DR GO; GO:0046530; P:photoreceptor cell differentiation; IMP.
DR GO; GO:0045449; P:regulation of transcription; IDA.
DR GO; GO:0007379; P:segment specification; IMP.
DR GO; GO:0008587; P:wing margin morphogenesis; IMP.
DR GO; GO:0016055; P:wnt receptor signaling pathway; IMP.
DR InterPro; IPR001606; ARID.
DR InterPro; IPR008938; ARM.
DR InterPro; IPR006031; XYPPX.
DR Pfam; PF01388; ARID; 1.
DR Pfam; PF02162; XYPPX; 8.
DR SMART; SM00501; BRIGHT; 1.
KW Transcription regulation; DNA-binding; Activator; Repressor;
KW Chromatin regulator; Nuclear protein; Developmental protein.
FT DOMAIN 997 1111
FT DOMAIN 1769 2517 EHD.
FT DOMAIN 19 1763 PRO-RICH.
FT DOMAIN 174 380 GLN-RICH.
FT DOMAIN 619 873 GLY-RICH.
FT DOMAIN 1222 1453 GLY-RICH.
FT DOMAIN 1271 1751 GLN-RICH.
FT DOMAIN 1730 1745 HIS-RICH.
FT DOMAIN 2589 2624 SER-RICH.
FT DOMAIN 2625 2716 ALA-RICH.
FT CONFLICT 61 61 MISSING (IN REF. 1).
FT CONFLICT 1169 1169 V -> G (IN REF. 1).
FT CONFLICT 1795 1795 M -> T (IN REF. 1).
FT CONFLICT 2637 2637 G -> E (IN REF. 1).
SQ SEQUENCE 2716 AA; 284063 MW; EFAE76CB51C7C675 CRC64;

Query Match 8.8%; Score 251.5; DB 1; Length 2716;
Best Local Similarity 23.9%; Pred. No. 2.1e-05;
Matches 152; Conservative 49; Mismatches 208; Indels 227; Gaps 36;

QY 3 PRNHPSR----PPSTSLPQGSAPLPSA-----PISSMPMPQYTMQYFVSQ----- 45
Db 292 PQLGPSQYRTPTPTNTSRGQSPYPPAHGQNSGSPSSPQQQQQQQQQQAGQPGGP 351
QY 46 -PHTLPP---LQPHHSQSPAPSHYMGQPPYRDLNRYPA--SSH-DVYASSAAPIMPHTT 98
Db 352 VPGGPPPGTGQPPQNTPTTSQSPYQRYPTTTPGLPAGGSNHRATYSTHQPENRPW 411
QY 99 VGLPPTSFSLSHPNPQAQAQAQSPHY-----PPHVSVL--PPASSAQSY-----P 142
Db 412 PGGSSSPSPGSGHPLPPA-----SPHHVBLQQQPPPPPHVSAAGPPSSSPGHAPSPSP 465
QY 143 QP--IAPAPRRDRRADFNGLPSGAFYSYDQKPGQWDPAANGAAPYPGKDSPTQVV-- 198
Db 466 QPSQASPSPHQELIGQNSDSSG-----GAHSGM-----GSGP-PGTNPQ-QVMRP 511
QY 199 -----GSQRRGILPSVPGR---ATPVTVNGVTGKNTTIPAKDADGKFPNCNKTYLH 250
Db 512 TPSPTGSSGSRMSMSPAQNHPISRPSAN-QSSSGGPMQQPPFVGAGGPPMP----- 562
QY 251 AKHLKRHLRHTGDRPYMCVLCCKTFRSDILKRHFQKCSIRRGNTGATHLSHPNHVK 310
Db 563 -----PH-----PGMPGGP-----PQQQ 575
QY 311 RSQQQAAAN-----PVKPVQDEVSVTPPPNGIPGTTY-----GEG 346
Db 576 QSQQQQAASNSASSASNSPQQTTPPPAPPENQGMNMTATPPPPQGAAGGYPMPPMHGGY 635
QY 347 AVNGNGLAPRPGYADHQTMGF-----PMSSVNGMG--RGQPE 382
Db 636 KMGGPGQSPGAQGYPPQPPQYPPGNYPRPQYPPGAYATGPPPPPTSQAGAGGANSMP 695
QY 383 DAFPGGRPHQGAP-----W-PQAPKQSPYLVPQGADPSGHQLNIDRNIEQVK--- 428

Db 696 GAQAGGYPGRMENHTGQYPPYQWVPPSPQQT-----VPGGAPGGAMVG-----NHVQKGTP 748

QY 429 -QPVVQDPKRPVMPGHP-----GHPGELDWTSMFQPPQAPGY-----MFSQ 468

Db 749 PPPVVGPPPPQSGSPRPLNYLKQHLKHKGYG-----SPTPPQPGQYNGPTGMHP 803

QY 469 SMPGGQEPH---AHVETERKYPTTTAGQESGMNG 501

Db 804 GMPMG-PPHMGPPHGPTNMG-PPTSTPPQSQMLQG 837

RESULT 3

BTD_DROME

ID_BTD_DROME STANDARD; PRT; 644 AA.

AC Q24266; Q9W319;

DT 15-JUL-1998 (Rel. 36, Created)

DT 15-JUL-1998 (Rel. 36, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE Transcription factor BTB (Butterfly protein).

GN BTB OR CG12653.

OS Drosophila melanogaster (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephydroidea; Drosophilidae; Drosophila.

OX NCBI_TaxID=7227;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=Canton-S; TISSUE=Embryo;

RX MEDLINE=94081952; PubMed=8259212;

RA Wimmer E.A., Jaekle H., Pfeifle C., Cohen S.M.;

RT "A Drosophila homologue of human Sp1 is a head-specific segmentation gene.";

RL Nature 366:690-694 (1993).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=Berkley;

RX MEDLINE=20196006; PubMed=10731132;

RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,

RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,

RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,

RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,

RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,

RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,

RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,

RA Balieu R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,

RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,

RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,

RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,

RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,

RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,

RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,

RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,

RA Fosler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,

RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,

RA Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,

RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,

RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,

RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,

RA Lasko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,

RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,

RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,

RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,

RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,

RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,

RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,

RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,

RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,

RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,

RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissenbach J.,

RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,

RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,

RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,

RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;

RT "The genome sequence of Drosophila melanogaster.";

RL Science 287:2185-2195 (2000).

CC -!- FUNCTION: REQUIRED FOR THE DEVELOPMENT OF THE ANTENNAL,

CC INTERCALARY AND MANDIBULAR SEGMENTS OF THE HEAD.

CC -!- SUBCELLULAR LOCATION: Nuclear (By similarity).

CC -!- DEVELOPMENTAL STAGE: FIRST EXPRESSED IN A STRIPE COVERING THE HEAD

CC ANLAGEN OF THE SYNCYTIAL BLASTODERM EMBRYO, PERSISTS THROUGH

CC GASTRULATION AND DECAYS DURING GERM BAND EXTENSION. EXPRESSED

CC LATER IN DEVELOPMENT IN A COMPLEX SPATIALLY RESTRICTED PATTERN.

CC -!- SIMILARITY: Contains 3 C2H2-type zinc fingers.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL; Z29361; CA82545.1; --.

DR EMBL; AE003448; AAF46518.1; --.

DR HSSP; P08047; 1SP2.

DR FlyBase; FBgn0000233; btd.

DR InterPro; IPR007087; Znf_C2H2.

DR Pfam; PF00096; zf-C2H2; 3.

DR ProDom; PD000003; Znf_C2H2; 1.

DR SMART; SM00355; Znf_C2H2; 3.

DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 3.

DR PROSITE; PS0157; ZINC_FINGER_C2H2_2; 3.

KW Transcription regulation; Activator; Zinc-finger; Metal-binding;

KW DNA-binding; Nuclear protein; Repeat.

FT ZN_FING 333 357 C2H2-TYPE 1.

FT ZN_FING 363 385 C2H2-TYPE 2.

FT ZN_FING 391 413 C2H2-TYPE 3.

FT DOMAIN 14 89 GLN-RICH.

FT DOMAIN 208 220 POLY-ALA.

FT DOMAIN 431 434 POLY-ALA.

FT DOMAIN 486 492 POLY-PRO.

FT DOMAIN 499 502 POLY-THR.

FT DOMAIN 515 519 POLY-SER.

FT DOMAIN 530 536 POLY-SER.

FT DOMAIN 596 599 POLY-SER.

SQ SEQUENCE 644 AA; 68581 MW; A0DB98C2AF938452 CRC64;

Query Match 8.3%; Score 236.5; DB 1; Length 644;

Best Local Similarity 26.1%; Pred. No. 2.9e-05;

Matches 111; Conservative 43; Mismatches 162; Indels 109; Gaps 20;

QY 8 SRPSTSLPQGSAPLPSAPISMPMPQYTMQPPVVSQPHLPLPQPHHSQSPAPHSYMG 67

Db 100 SAPPSLSSGSSGSSGSSPLYGKP----PMKLELPYQASSTGTASPNSSIQASPSASV 155

QY 68 QPPYRPDLNRYPASSHDVYAS-SAAPIMPHTTVGSLPPTSFL-----SHNPQ-- 114

Db 156 SPSIFPS----PAQS---FASISASPTTTTIA--PPTTAAAGALAGSPTSSSPSSSAA 206

QY 115 -----AQQAQSQSPHYPPPHSVLPPPASSAQSPQPIAPAPPRDRADF 157

Db 207 SAAAAAALGAAVASAAYGWNTAYSGLGPARSQFPYAQ-----YASDY 257

QY 158 NN---GLFSGA--FSYSDGKPGQWDPVAANGAAPPYGDSPRTQVVGSGRRGILPSVPG 212

Db 258 YGNAVGMSSAAWFHQERLYQPW-----SSQSYPGFNFDIAFTQLQRRSVRCTCPN 311

QY 213 RATPVTNGVNGTKNTTIPAKDADGK-----FPCPNCKNTYLAHKLKRHLRHTGDRPYM 268

Db 312 ----CTNEMSGL---PPIVGPDERGRKHQICHIPGICERLYGKASHLKTFLRWHTGERPFL 364

QY 269 CVLCKDTFSRSDILKRHFQ-----KCSIRRGNTGATHLS-HPNAHVK----- 310

Db 365 CLTCGKRFSRSDILKRHFQ-----KCSIRRGNTGATHLS-HPNAHVK----- 310

QY 311 -RSQQAAA-----NPVKP---VQDEVSVSTVPppNGIP-GTTYGEGAVNGNL 353
 :::||||| : | | | | |
Db 425 AEAKEQAAAAIAKLEKKKSGKPLTPVEFKQEQPDTPLVNYAPYANLYOHTSAGSSV 484

Qy 354 APARF 358
Db 485 NPPPF 489

RESULT 4

DRPL HUMAN STANDARD; PRT; 1185 AA.

ID DRPL HUMAN STANDARD; PRT; 1185 AA.

AC P54259; Q99495; Q99621; Q9UEK7;

DT 01-OCT-1996 (Rel. 34, Created)

DT 10-OCT-2003 (Rel. 42, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE Atrophin-1 (Dentatorubral-pallidoluysian atrophy protein).

GN DRPLA.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Brain, and Cerebellum;

RX MEDLINE=95144175; PubMed=7842016;

RA Nagafuchi S., Yanagisawa H., Ohsaki E., Shirayama T., Tadokoro K.,

RA Inoue T., Yamada M.;

RT "Structure and expression of the gene responsible for the triplet

RT repeat disorder, dentatorubral and pallidoluysian atrophy (DRPLA).";

RL Nat. Genet. 8:177-182(1994).

RN [2]

RP SEQUENCE FROM N.A.

RC TISSUE=Brain;

RX MEDLINE=96026098; PubMed=7485154;

RA Onodera O., Oyake M., Takano H., Ikeuchi T., Igarashi S., Tsuji S.;

RT "Molecular cloning of a full-length cDNA for dentatorubral-

RT pallidoluysian atrophy and regional expressions of the expand alleles

RT in the CNS.";

RL Am. J. Hum. Genet. 57:1050-1060(1995).

RN [3]

RP SEQUENCE FROM N.A.

RC MEDLINE=96262314; PubMed=8955642;

RA Margolis R.L., Li S.-H., Young W.S., Wagster M.V., Stine O.C.,

RA Kidwai A.S., Ashworth R.G., Ross C.A.;

RT "DRPLA gene (atrophin-1) sequence and mRNA expression in human

RT brain.";

RL Brain Res. Mol. Brain Res. 36:219-226(1996).

RN [4]

RP SEQUENCE FROM N.A., AND POLYMORPHISM OF POLY-GLN REGION.

RC TISSUE=Brain;

RX MEDLINE=97228904; PubMed=9074930;

RA Ansari-Lari M.A., Shen Y., Muzny D.M., Lee W., Gibbs R.A.;

RT "Large-scale sequencing in human chromosome 12p13: experimental and

RT computational gene structure determination.";

RL Genome Res. 7:268-280(1997).

RN [5]

RP SEQUENCE OF 1-76 FROM N.A.

RX MEDLINE=97005364; PubMed=8852663;

RA Yanagisawa H., Fujii K., Nagafuchi S., Nakahori Y., Nakagome Y.,

RA Akane A., Nakamura M., Sano A., Komure O., Kondo I., Jin D.K.,

RA Soerensen S.A., Potter N.T., Young S.R., Nakamura K., Nukina N.,

RA Nagao Y., Tadokoro K., Okuyama T., Miyashita T., Inoue T.,

RA Kanazawa I., Yamada M.;

RT "A unique origin and multistep process for the generation of expanded

RT DRPLA triplet repeats.";

RL Hum. Mol. Genet. 5:373-379(1996).

RN [6]

RP SEQUENCE OF 470-725 FROM N.A.

RC TISSUE=Brain cortex;

RX MEDLINE=93315145; PubMed=8325628;

RA Li S.-H., McInnis M.G., Margolis R.L., Antonarakis S.E., Ross C.A.;

RT "Novel triplet repeat containing genes in human brain: cloning,

SQ	SEQUENCE	1185 AA; 124773 MW; 5135744CEE491C18	CRC64;
	Query Match	8.2%; Score 234.5; DB 1; Length 1185;	
	Best Local Similarity	22.0%; Pred. No. 6.5e-05;	
	Matches 123; Conservative	46; Mismatches 193; Indels 197; Gaps 23;	
QY	7	PSRPPSTSLPQGSAPLPSAPISSMPMPQYTMQPYVSPQHTLPLPQPHHSQSPAPHSYM	66
Db	272	PTKPTT--PVGGNLPAP-----PPANFPHVTNLP-----	303
QY	67	GPPYRPDLNRYPASSHDVYASSAAPIMPHTVVGLPPTSELSHPNPQAQA	119
Db	304	--PALRPLNNASASPPGL---GAQLPLGHL-----PSPHAMGQGMGGLPPG	345
QY	120	-QSPHY-PPPHSVLPPASSAQSYQPIAPAPRRDRADFNGLPSGAFSYSDGKPGQWD	177
Db	346	PEKGPTLAPSPHS-LPPASS-----APAPPMRFYSSSSSSSSSSSSSSSSSS	397
QY	178	PVAANGAAP-YPGKDSPTQV-VGSQRRGILPSVPGRATPVTNGVNGTGKNTTIPAKDA	235
Db	398	FPASQALPSYHSPPTSLSVSNQPKYQTPSLPSQAV-----WS	439
QY	236	DGKFCPCNCKNTYLHAKHLKRHLRHTGDRPYMCVLCCKDTFSRSDILKRHFQKCSIRGN	295
Db	440	QGPPPPPPYGRLLANSN-----AHGPPF-----P	464
QY	296	PTGATHLSHP---NAHVKRSQQQAAANPVKPVQDEVSVTPPPNGIPGTTYGEGAVNGN-	351
Db	465	STGAQSTAHPPVSTHHHHQQQQQQQQQQQHGNSGPPPPGAPHPHLEGSSSHAHP	524
QY	352	-----GLAPARPG-----YADHQTMGFPMSVNGMGRGQPEDAFPGRPH	391
Db	525	YAMSPSLGLRPPYPGPAHLPPPHSVSVSQAGPNPVPVSSSSSSSSSSSSSSS	584
QY	392	-----QGAPWQAP-----KQSPYLVPQGADPSGHQLNIDRNIQVQPVVQPKR	437
Db	585	PSQPGQGAPYPPFPVPTVTSSATLSTVIATVASSPAGY-----KT	625
QY	438	PVMPGHPGHPGELDWTSMFQPAEGYMFQSMPGGQEPHIAHVETERKYPTTTAGQES	497
Db	626	ASPPGPPPYGKRAPSPGAYKTATPPGY-----KPGSP-----PSFRTGTTP	666
QY	498	GMNGLYLASTMSGDGTVP	516
Db	667	GYRG---TSPPAGPGTFKP	682
RESULT 5			
PRD1_MOUSE	STANDARD;	PRT;	856 AA.
ID PRD1_MOUSE			
AC Q60636;			
DT 28-FEB-2003 (Rel. 41, Created)			
DT 28-FEB-2003 (Rel. 41, Last sequence update)			
DT 10-OCT-2003 (Rel. 42, Last annotation update)			
DE PR-domain zinc finger protein 1 (Beta-interferon gene positive-regulatory domain 1 binding factor) (BLIMP-1).			
GN PRD1 OR BLIMP1.			
OS Mus musculus (Mouse).			
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
OX NCBI_TaxID=10090;			
RN [1]			
RP SEQUENCE FROM N.A.			
RC STRAIN=BALB/c;			
RX MEDLINE=94221646; PubMed=8168136;			
RA Turner C.A., Mack D.H., Davis M.M.;			
RT "Blimp-1, a novel zinc finger-containing protein that can drive the			
RT maturation of B lymphocytes into immunoglobulin-secreting cells.;"			
RL Cell 77:297-306(1994).			
RN [2]			
RP SEQUENCE FROM N.A.			
RC STRAIN=129;			
RX MEDLINE=20572522; PubMed=11121475;			
RA	Tunyaplin C., Shapiro M.A., Calame K.L.;		
RT "Characterization of the B lymphocyte-induced maturation protein-1			
RT (Blimp-1) gene, mRNA isoforms and basal promoter.;"			
RL Nucleic Acids Res. 28:4846-4855(2000).			
CC -!- FUNCTION: Transcriptional repressor that binds specifically to the			
CC PRDI element in the promoter of the beta-interferon gene. Drives			
CC the maturation of B lymphocytes into Ig secreting cells.			
CC -!- SUBCELLULAR LOCATION: Nuclear (By similarity).			
CC -!- SIMILARITY: Contains 1 SET domain.			
CC			
CC This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC the European Bioinformatics Institute. There are no restrictions on its			
CC use by non-profit institutions as long as its content is in no way			
CC modified and this statement is not removed. Usage by and for commercial			
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC or send an email to license@isb-sib.ch).			
CC			
DR EMBL; U08185; AAA19252.1; -.			
DR EMBL; AF305539; AAG42212.1; -.			
DR EMBL; AF305534; AAG42212.1; JOINED.			
DR EMBL; AF305535; AAG42212.1; JOINED.			
DR EMBL; AF305536; AAG42212.1; JOINED.			
DR EMBL; AF305537; AAG42212.1; JOINED.			
DR EMBL; AF305538; AAG42212.1; JOINED.			
DR PIR; A53503; A53503.			
DR HSSP; P08048; 7ZNF.			
DR TRANSFAC; T02316; -.			
DR MGD; MGI:99655; Prdm1.			
DR InterPro; IPR001214; SET.			
DR InterPro; IPR007087; Znf_C2H2.			
DR Pfam; PF00856; SET; 1.			
DR Pfam; PF00096; zf-C2H2; 5.			
DR ProDom; PD000003; Znf_C2H2; 1.			
DR SMART; SM00317; SET; 1.			
DR SMART; SM00355; Znf_C2H2; 4.			
DR PROSITE; PS50280; SET; 1.			
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 4.			
DR PROSITE; PS01157; ZINC_FINGER_C2H2_2; 4.			
KW Transcription regulation; Repressor; DNA-binding; Zinc-finger;			
KW Metal-binding; Nuclear protein; Repeat.			
FT DOMAIN 118 237 SET.			
FT ZN_FING 606 628 C2H2-TYPE.			
FT ZN_FING 634 656 C2H2-TYPE.			
FT ZN_FING 662 684 C2H2-TYPE.			
FT ZN_FING 690 712 C2H2-TYPE.			
SQ SEQUENCE 856 AA; 95835 MW; B9AC6FC2E29ECF4A			CRC64;
Query Match			
Best Local Similarity 8.1%; Score 233; DB 1; Length 856;			
Matches 92; Conservative 40; Mismatches 160; Indels 86; Gaps 14;			
QY	13	TSLPQGSAPLPSAPISSMPMPQYTMQPYVSPQ-HTLPLQPHHSQSPAPHSYMGQPPY	71
Db	357	THSPLPSSTTPSPASS--SPEQSLKSSSPHSSPGNTVSPPLAPGLPEHRDSYSLNVSYG	414
QY	72	RPDLNRYPASSHDVYASSAAPIMPHTTVGS-----LPPTSFLSH-----PNPQAQA	117
Db	415	SEGLGSYPG-----YA--PAPHLPPAFIPSYNAHYPKFLPPYGISSNGLSTMNNGIN	467
QY	118	QAQOSPHYPP-----PHSVLPPASSAQSY-----QPIAPAPRRDRRADFNNGL	161
Db	468	NFSLFRLYPVYSNLLSGSLPHPMNLNPASLPSSLPDGGARRLLPPEHPKEVLIP----	523
QY	162	PSGAFSYSDGKPGQWDVPAANGAAPYPGKDSPTQVVGSGQRRGIL--PSVPGRATPVTN	219
Db	524	PHSAFSLTGAAASMKDESSPSPGSPTAGTAATSEHVQPKATSSVMAAPSTDGAMNLKN	583
QY	220	GVNGTG-KNTTIPAKDADGK-----FPCPCNCKTY 248	
Db	584	KRNMTGYKTLPLYPLKKQNGKIKYECNVCAKTFGQLSNLKVHLRVHSGERPFKQTCNKG	643
QY	249	LHAKHLKRHLRHTGDRPYMCVLCCKDTFSRSDILKRHFQ-----KCSIRGNPTGAT	300

RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Datchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RL human and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [4]
RP SEQUENCE OF 339-417 FROM N.A.
RX MEDLINE=88216608; PubMed=2452975;
RA Chavrier P., Lemaire P., Revelant O., Bravo R., Charnay P.;
RT "Characterization of a mouse multigene family that encodes zinc
RT finger structures.";
RL Mol. Cell. Biol. 8:1319-1326(1988).
RN [5]
RP FUNCTION.
RX MEDLINE=90214631; PubMed=1969796;
RA Chavrier P., Vesque C., Galliot B., Vigneron M., Dolle P., Duboule D.,
RA Charnay P.;
RT "The segment-specific gene Krox-20 encodes a transcription factor
RT with binding sites in the promoter region of the Hox-1.4 gene.";
RL EMBO J. 9:1209-1218(1990).
RN [6]
RP FUNCTION.
RX MEDLINE=91234777; PubMed=1674431;
RA Gilardi P., Schneider-Maunoury S., Charnay P.;
RT "Krox-20: a candidate gene for the regulation of pattern formation in
RT the hindbrain.";
RL Biochimie 73:85-91(1991).
RN [7]
RP DOMAINS.
RX MEDLINE=92285130; PubMed=1598206;
RA Vesque C., Charnay P.;
RT "Mapping functional regions of the segment-specific transcription
RT factor Krox-20.";
RL Nucleic Acids Res. 20:2485-2492(1992).
CC -!- FUNCTION: SEQUENCE SPECIFIC DNA-BINDING TRANSCRIPTION FACTOR.
CC BINDS TO TWO SPECIFIC DNA SITES LOCATED IN THE PROMOTER REGION OF
CC HOX-1.4. MAY PLAY A ROLE IN THE REGULATION OF HINDBRAIN
CC SEGMENTATION, MIGHT ACT IN COMBINATION WITH THE HOX NETWORK TO
CC SPECIFY ODD AND EVEN RHOMBOMERES, AND MIGHT PARTICIPATE IN THE
CC CONTROL OF THE EXPRESSION OF SOME OF THE HOMEOBOX CONTAINING
CC GENES.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=Long;
CC IsoId=P08152-1; Sequence=Displayed;
CC Name=Short;
CC IsoId=P08152-2; Sequence=VSP_006864;
CC TISSUE SPECIFICITY: EXPRESSED MAINLY IN ADULT THYMUS AND EMBRYONIC
CC NERVOUS SYSTEM.
CC -!- INDUCTION: ACTIVATED DURING G0/G1 TRANSITION IN CULTURED CELLS.
CC -!- SIMILARITY: BELONGS TO THE EGR FAMILY OF C2H2-TYPE ZINC-FINGER
CC PROTEINS.
CC -!- SIMILARITY: Contains 3 C2H2-type zinc fingers.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M24377; AAA39379.1; -.
DR EMBL; M24376; AAA39379.1; JOINED.
DR EMBL; M24377; AAA39380.1; -.
DR EMBL; M24376; AAA39380.1; JOINED.
DR EMBL; X06746; CAA29921.1; -.
DR EMBL; BC009093; AAH09093.1; -.
DR EMBL; M20759; AAA39381.1; -.
DR PIR; A30136; A30136.
DR PIR; S00256; S00256.
DR HSSP; P08046; IAAI.
DR TRANSFAC; T00454; -.
DR MGD; MGI:95296; Egr2.
DR InterPro; IPR007087; Znf_C2H2.
DR Pfam; PF000096; zf-C2H2; 3.
DR ProDom; PD000003; Znf_C2H2; 2.
DR SMART; SM00355; Znf_C2H2; 3.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 3.
DR PROSITE; PS0157; ZINC_FINGER_C2H2_2; 3.
KW Transcription regulation; Activator; DNA-binding; Nuclear protein;
KW Repeat; Zinc-finger; Metal-binding; Alternative splicing.
FT DOMAIN 167 173 POLY-PRO.
FT ZN_FING 337 361 C2H2-TYPE 1.
FT ZN_FING 367 389 C2H2-TYPE 2.
FT ZN_FING 395 417 C2H2-TYPE 3.
FT VARSPIC 1 50 Missing (in isoform Short).
FT /FTId=VSP_006864.
SQ SEQUENCE 470 AA; 49819 MW; 6771273C77960F0 CRC64;

Query Match 7.8%; Score 223; DB 1; Length 470;
Best Local Similarity 22.0%; Pred. No. 0.0001;
Matches 95; Conservative 49; Mismatches 156; Indels 132; Gaps 17;

QY 9 RPPSTSLPQGSAPLPSAPISS--MPMPQYTMQPOYP-----VSQP 46
Db 69 RPLDLPYSSFAPY-SAPRNQFTTYMGKFSIDPQYPGASCYPEGIIVSAGILQGVTPP 127
QY 47 HTLPPLQPHHSQSAP-----HSYMGQPPYPPDLNRYPASSHDVYA 87
Db 128 ASTTASSSVTSASPNPLATGTLGVCVTMSQTQPELDHLYSPPPPPP-----YSGCTGDLQ 183
QY 88 SSAAPIMPHTTVGS-----LPPTSLFLSHNPQAQAQAQSPHY-----PPHSVLP 133
Db 184 DPSAFLSPFSTTSTSSLAYQPPPSYSPSPKAMPDGLIPMDYPGFFPSPCQDRPHGAAG 243
QY 134 PASSAQSYQPI-----APAPRRDRADFNNGLPS-----GAFSYSDGKPGQWDFVAA 181
Db 244 P--DRKPPCPDLSLRVPPPLTLPLSTIRNFTLGGPGAGVTGPGASGGGEGPRLPGSGSAA 301
QY 182 NGAAPYPKDSPTQVVGSGQRRGILPSVPGRATPTVNGVNGTGKNTTIPAKDADGKFP 241
Db 302 VTATPYNPHHLPLRPIL---RPRKYPNRPSPK-TPV-----HERPYPC 339
QY 242 P--NCNKTYLHAKHLKRHLRLHRTGDRPYMCVLCQDTFSRSDILKRHFQKCSIRRGNTGA 299
Db 340 PAEGCDRFRSDELTRHIRHTGHPFQCRICMRNFRSRSDLTHIR----- 387
QY 300 THLS-----HPNAHVRSQQQAANPKVPQVQDEVSVTPPPNGI 338
Db 388 THTGKFPACDYCGRKFAFSDEKRTKIHLRQERKSSA-PSAPPSAQSSASGPGGSQA 446
QY 339 PGITYGEGAVNG 350
Db 447 GGSILCGNSAIGG 458

RESULT 8
YQ36 CAEEL
ID YQ36 CAEEL STANDARD; PRT; 963 AA.
AC Q09457;

DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Putative cuticle collagen C09G5.6.
GN C09G5.6.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Palmer S.;
RL Submitted (NOV-1994) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Nematode cuticles are composed largely of collagen-like
CC proteins. The cuticle functions both as an exoskeleton and as a
CC barrier to protect the worm from its environment (By similarity).
CC -!- SUBUNIT: Collagen polypeptide chains are complexed within the
CC cuticle by disulfide bonds and other types of covalent cross-links
CC (By similarity).
CC -!- SIMILARITY: Belongs to the cuticular collagen family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; Z46791; CAA86755.1; -.
DR PIR; T19140; T19140.
DR WormPep; C09G5.6; CE01486.
DR InterPro; IPR008161; Clg_helix.
DR InterPro; IPR002486; Col_cuticle_N.
DR InterPro; IPR008160; Collagen.
DR Pfam; PF01484; Col_cuticle_N; 1.
DR Pfam; PF01391; Collagen; 3.
DR ProDom; PD000007; Clg_helix; 2.
KW Hypothetical protein; Cuticle; Connective tissue; Repeat;
KW Multigene family; Collagen.
FT DOMAIN 392 423 TRIPLE-HELICAL REGION.
FT DOMAIN 441 503 TRIPLE-HELICAL REGION.
FT DOMAIN 506 567 TRIPLE-HELICAL REGION.
FT DOMAIN 663 666 POLY-PRO.
FT DOMAIN 685 688 POLY-PRO.
SQ SEQUENCE 963 AA; 107031 MW; AFF895A75909F66E CRC64;

Query Match 7.7%; Score 222; DB 1; Length 963;
Best Local Similarity 21.4%; Pred. No. 0.00022;
Matches 145; Conservative 56; Mismatches 205; Indels 272; Gaps 36;

QY 7 PSRPPSTSLPQGS-----AELPSAPIS-----SMMPQYTMQPY----- 41
Db 191 PQQPPSTSAPHSSPNKRTSLYNPQPPTKGYPTNPRVYNPPQPNYTRQPTYPEDNRAPY 250

QY 42 -PVSQPHTLPLQPHHSQSPAPHSYMGQ-PPYRPDL--NRYPASSHDVYASSAAPIMPHT 97
Db 251 KPTRSPNTPPRPQ------SGGYDSGQTPTPSSPRIYNTRRPNNHGPGYPEDQVPTAPPV 305

QY 98 T-VGSLPPTSLFLSHPNPQAQAQSQSPHY-PEHS-----VLPPASSAQSYFQ----- 143
Db 306 PGQQRVPPTQTRNPPNPTNRQ----PSRPVPTSDGHIEATTPYNPSAQYPTGKRGSH 361

QY 144 ---PIAPAP-----PRDRRADFNNGLPFG-----AFSYSDG----- 171
Db 362 GFGEQRPRTGRPRGNPCDQCSAQPNHCPSPGPGRGRPGPPGFGQDGRGLRGLNGGY 421

QY 172 ---KPQGWDPV-----AANGAAPYPKGD----- 191
Db 422 SGVQPSYDYPVIGCVQCPIGPPGERGPDGTPGVGEDGIDGEGVNGQDGPAPGAPGY 481
QY 192 -----SPRT-----QVVGSGQRRRGLPLPSVPGR-ATPVTNGVNGTGKNTTI 230

Db 482 HGMNGSPGTPGKPLPGRNQGSCSKSIPGPPGQPGVM-GVFGRDGDPGTDGEGHQDGSPI 540
QY 231 ---PAKD-----ADGKF-PCPNCKNTY-----LHAKHLKRHLRLRHTGD 264
Db 541 QGPPGRDGTSGPDGQPGVSAPGAPGTDGGYCPCPKRSSKFDNFNDAYNDDEKRGLEHR-P 599
QY 265 RPYMCVLCKDTFSRSDILKRHFQKCSIRRGNTGATHLSHPNAHVXRSQQAAANPVKPV 324
Db 600 RGYDSERAEPRPRQTVRTNTYDENS-----GAHQRRPNY----- 635
QY 325 QDEVSSTVPPPNIGPCTTYGEGAVNGNGLAPARPGYADHQTMGFPMSSVNGMGRGQPEDA 384
Db 636 --EPSAEVAPPR-----QDRYDEERVREP-----PPKRP 663
QY 385 FPGGRPHQGAPWPQAPKQSPYLVQPGADPSGHQLNIDRNIEQV----- 427
Db 664 PP---PHRQTPHELYPEEQPYVRRPPPPQNRGNYEVSREYVPEAPRPRQGYEHSSGYG 720
QY 428 -----KQVVVQDKRPVMPGHPGHPGELJWTSMFQ-----QAPE-GY-----MFSQS 469
Db 721 GDDRKEQPKYME-SRPV--DEPRY--ETDAPSRPLKKVELIRHPERGYDRRQPSYEDS 775
QY 470 MPGGQEPIHAHVETERY 487
Db 776 KPRQEPRRRYETEAPRY 793

RESULT 9
SP7_HUMAN STANDARD; PRT; 431 AA.
AC Q8TDD2;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Transcription factor Sp7 (Zinc finger protein osterix).
GN SP7 OR OSX.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Osteosarcoma;
RA Ganss B.W.;
RT "cDNA sequence, gene structure and chromosomal localization of the
RT human osterix (OSX) gene."
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Nakashima K., Zhou X., de Crombrughe B.;
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RX PubMed=14604442;
RA Milona M.-A., Gough J.E., Edgar A.J.;
RT "Expression of alternatively spliced isoforms of human Sp7 in
RT osteoblast-like cells."
CC BMC Genomics 4:43-43(2003).
CC -!- FUNCTION: Transcriptional activator essential for osteoblast
CC differentiation. Binds to Sp1 and EKLF consensus sequences and to
CC other G/C-rich sequences (By similarity).
CC -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC -!- SIMILARITY: BELONGS TO THE SP1 FAMILY OF C2H2-TYPE ZINC-FINGER
CC PROTEINS.
CC -!- SIMILARITY: Contains 3 C2H2-type zinc fingers.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

```
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF477981; AAL84281.1; -.
DR EMBL; AF466179; AAO33377.1; -.
DR EMBL; AY150673; AAN85556.1; -.
DR Genew; HGNC:17321; SP7.
DR MIM; 606633; -.
DR InterPro; IPR007087; Znf_C2H2.
DR Pfam; PF00096; zf-C2H2; 3.
DR ProDom; PD000003; Znf_C2H2; 1.
DR SMART; SM00355; Znf_C2H2; 3.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 3.
DR PROSITE; PS50157; ZINC_FINGER_C2H2_2; 3.
KW Transcription regulation; Zinc-finger; Metal-binding; DNA-binding;
KW Nuclear protein; Repeat; Activator.
FT ZN_FING 294 318 C2H2-TYPE 1.
FT ZN_FING 324 348 C2H2-TYPE 2.
FT ZN_FING 354 376 C2H2-TYPE 3.
SQ SEQUENCE 431 AA; 44994 MW; 454A6FEA84309FF9 CRC64;

Query Match
Best Local Similarity 7.7%; Score 221.5; DB 1; Length 431;
Matches 111; Conservative 30; Mismatches 139; Indels 181; Gaps 22;

CC 11 PSTSL-----PQSAPLPSA-----PIS-SMPMPQYTMQPYVSPHTLPLPQPHS 57
CC ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
CC 67 PFTSTNGLLSAGSPAPTSGYANDYPPFSHSFPPTGTQDPGLLVPKGHS-----S 118

CC 58 QSPAPHSYMGPPYRDLNRYPASSSHDVYASSAAPIMPTTVGSLPPTSFLSHP----- 111
CC ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
CC 119 SDCLPSVYTSMDTHPYGSWKAGIH-----AGISGP--GNTPTPWDMHPGGNWLG 169

CC 112 NPQAQAQAQ-----SPHYPPHSVLPPASSAQSYQPQIAPAPRRDRADFNGLPSGAF 166
CC ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
CC 170 GGQGGDGLQGLTGPAPPLNPQLP-----TYPDFAPLNP----- 207

CC 167 SYSDGKPGQWDVAANGAAPYPG-----KDSPTQVVSQGR---RGILPSV 210
CC ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
CC 208 -----APYPAPHLLQPGQHVLPQDVYKPKAVGNSGQLEGSGGAKPP 249

CC 211 PGRATPVTVNGVTGKNTTIPAKDADGKFFC-----GRSSCDPCNCOELERLGAAGLRKKPIHSCHIP 242
CC ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
CC 250 RGASTGGSGGYGSGA-----GRSSCDPCNCOELERLGAAGLRKKPIHSCHIP 299

CC 243 NCNKTYLHAKHLKRHLRHTGDRPYMC--VLCKDTFSRSDILKRHFQ-----KCSIR 292
CC ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
CC 300 GCGKVGKASHLKAHLRHWHTGERPFVNCWLPFCGKFRTRSDELERHVRTHREKFTCLLC 359

CC 293 RGNPTGATHLS-HPNAHVKRSQQQAANPVKPVQDEVSVTVPPNGIPGTTYGEGAVNGN 351
CC ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
CC 360 SKFRTRSDDLHLSKHQTHGE-----PGPGPPPSG--PKELGEGRSTGE 399

CC 352 GLAPARPGYADHQTMGFFPMSSVNGMGRGQPEDAFPPGRPHQ 392
CC ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
CC 400 HEASQTPR-----PSAS-----PATPEKA-PGGSPEQ 425

RESULT 10
YANB_SCHPO STANDARD; PRT; 582 AA.
ID YANB_SCHPO
AC Q10076;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical zinc finger protein C3H1.11 in chromosome I.
GN SPAC3H1.11.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
```


Db 314 PSNCSPSVKSSGSPSYHAEQEVNVNSYNGGIPSTSYNDTPQSVTGSYNSGETMSTYLNQ 373

QY 179 VAANGAAPYPGKDSPTQTQVVGSGRRGILPSVPGRATPVTNGVNGTGKNTTIPAKDADGK 238

Db 374 TNTSGRSP---NSMEATEQIGTIGTDGSMKRRKR-----QPSNRKTSVPRSPGGKS 422

QY 239 FPCPNCNKTYLHAKHLKRHLR-HTGDRPYMCVLCKDTFSRSDILKRHFQKCSIRRGNPT 297

Db 423 FVCDECSKKFKRSEHLRRHIRSLHTSEKPFVCI-CGRFSRRDNLQH-----ERLHVN 475

QY 298 GATHLS----HPNAHVKRSQQAAANPVKPVQDEVSSTVPPNGIPGT----- 341

Db 476 ASPLACFFQPSGYSSGAPGVQPQKPIED--LNKIPINQGMDSQIENTNMLSSQR 533

QY 342 -----TYGEGAVNGNGLAPARPGYADHQTMGFPMSSVNGMGRGQPEDAFPGGRPHQAPW 396

Db 534 PLSQIVPEIAAYPNISRPELLSKLPVQTPNQKMLNPM-----HQYQPY 579

QY 397 PQA 399

Db 580 PSS 582

RESULT 11

EGR2_RAT

ID EGR2_RAT STANDARD; PRT; 470 AA.

AC P51774;

DT 01-OCT-1996 (Rel. 34, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE Early growth response protein 2 (EGR-2) (Krox-20 protein).

GN EGR2 OR EGR-2 OR KROX-20 OR KROX20.

OS Rattus norvegicus (Rat).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

OX NCBI_TaxID=10116;

RN [1]

RP SEQUENCE FROM N.A. (ISOFORM LONG).

RC STRAIN=Sprague-Dawley; TISSUE=Brain;

RA Zaharic T., Durtschi B.A., Mason-Parker S.E., Abraham W.C.,

RA Tate W.P.;

RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A. (ISOFORM SHORT).

RC STRAIN=Wistar; TISSUE=Hippocampus;

RX MEDLINE=96192065; PubMed=8619872;

RA Inokuchi K., Murayama A., Ozawa F.;

RT "mRNA differential display reveals Krox-20 as a neural plasticity-

RT regulated gene in the rat hippocampus.";

RL Biochem. Biophys. Res. Commun. 221:430-436(1996).

CC -!- FUNCTION: SEQUENCE SPECIFIC DNA-BINDING TRANSCRIPTION FACTOR.

CC BINDS TO TWO SPECIFIC DNA SITES LOCATED IN THE PROMOTER REGION OF

CC HOX-1.4 (BY SIMILARITY).

CC -!- SUBCELLULAR LOCATION: Nuclear.

CC -!- ALTERNATIVE PRODUCTS:

CC Event=Alternative splicing; Named isoforms=2;

CC Name=Long;

CC IsoId=P51774-1; Sequence=Displayed;

CC Name=Short;

CC IsoId=P51774-2; Sequence=VSP 006865;

CC -!- SIMILARITY: BELONGS TO THE EGR_FAMILY OF C2H2-TYPE ZINC-FINGER

CC PROTEINS.

CC -!- SIMILARITY: Contains 3 C2H2-type zinc fingers.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL; U78102; AAB36783.1; -.

DR EMBL; D83508; BAA11932.1; -.

DR PIR; JC4716; JC4716.

DR HSSP; P08046; LAAY.

DR InterPro; IPR007087; Znf_C2H2.

DR Pfam; PF00096; zf-C2H2; 3.

DR ProDom; PD000003; Znf_C2H2; 2.

DR SMART; SM00355; Znf_C2H2; 3.

DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 3.

DR PROSITE; PS0157; ZINC_FINGER_C2H2_2; 3.

DR Transcription regulation; Activator; DNA-binding; Nuclear protein;

KW Repeat; Zinc-finger; Metal-binding; Alternative splicing.

KW DOMAIN 167 173 POLY-PRO.

FT ZN_FING 337 361 C2H2-TYPE 1.

FT ZN_FING 367 389 C2H2-TYPE 2.

FT ZN_FING 395 417 C2H2-TYPE 3.

FT VARSP LIC 1 50 Missing (in isoform Short).

FT /FTid=VSP 006865.

SQ SEQUENCE 470 AA; 49849 MW; 2B28185752A91F76 CRC64;

Query Match 7.7%; Score 221; DB 1; Length 470;

Best Local Similarity 22.5%; Pred. No. 0.00013;

Matches 97; Conservative 48; Mismatches 155; Indels 132; Gaps 18;

QY 9 RPPSTSLPQGSAPLPSAPISS--MPMPQYTMQQYP-----VSQP 46

Db 69 RPLDLPYPSSFAPI-SAPRNQTFTYMGKFSIDPQYFGASCYPEGIINIVSAGILQGVTP 127

QY 47 HTLPPLQPHHSQSPAP-----HSYMGQPPYRPDLNRYPASSHDVYA 87

Db 128 ASTTASSSVTSASFNPLATGLGVCTMSQTQPELDHLYSPPPPPP-----YSGCTGDL YQ 183

QY 88 SSAAPIM--PHTTVGSL---PPTSFLSHPNPQAQAQAQQSPHY P-----PPHSVLP 133

Db 184 DPSAFLSPPTTSTSSLAYQPPSPSPKPMDPGLIPMIPDYPGFFPSPCQRDPHGAAAG 243

QY 134 PASSAQSYQPPI-----APAPRRDRRADFNNGLP S-----GAFSYSDGKPGQGWDPVAA 181

Db 244 P--DRKFPFPCPLDSL RVPPPLTPLSTIRNFTLGGPSAGVTGPGASGGEGPRLPGSGSAA 301

QY 182 NGAAPYPGKDSRPTQVVGSQRRGILPSVPGRATPVTNGVNGTGKNTTIPAKDADGKFP C 241

Db 302 VTATPYNPHLPLRPIL----RPRKYPNRPSK-TPV-----HERPYPC 339

QY 242 P--NCNKTYLHAKHLKRHLRLRHTGDRPYMCVLCKDTFSRSDILKRHFQKCSIRRGNTGA 299

Db 340 PAEGCDRRFSRDELTRHIRIHTGHKPFQCRICMRNFSRSDHLTTHIR----- 387

QY 300 THLS-----HPNAHVKRSQQAAANPVKPVQDEVSSTVPPNNGI 338

Db 388 THTGEKPFACDYCGRKFA RSDERKRHTKIHLRQKERKSSA-PSSSASQA SSGPGGSQA 446

QY 339 PGTTYGEGAVNG 350

Db 447 GGS LCGNSAIGG 458

RESULT 12

MLL2_HUMAN

ID MLL2_HUMAN STANDARD; PRT; 5262 AA.

AC O14686; O14687;

DT 10-OCT-2003 (Rel. 42, Created)

DT 10-OCT-2003 (Rel. 42, Last sequence update)

DT 15-MAR-2004 (Rel. 43, Last annotation update)

DE Myeloid/lymphoid or mixed-lineage leukemia protein 2 (ALL1-related

DE protein).

GN MLL2 OR ALLR.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A. (ISOFORMS 1; 2 AND 3).

RX MEDLINE=97388474; PubMed=9247308;

RA Prasad R., Zhadanov A.B., Sedkov Y., Bullrich F., Druck T.,
RA Rallapalli R., Yano T., Alder H., Croce C.M., Huebner K., Mazo A.,
RA Canaani E.;
RT "Structure and expression pattern of human ALR, a novel gene with
RT strong homology to ALL-1 involved in acute leukemia and to Drosophila
RT trithorax.";
RL Oncogene 15:549-560(1997).
RN [2]
RP INTERACTION WITH ASC-2/NCOA6 CONTAINING COMPLEX.
RC TISSUE=Cervical carcinoma;
RX MEDLINE=22371496; PubMed=12482968;
RA Goo Y.-H., Sohn Y.C., Kim D.-H., Kim S.-W., Kang M.-J., Jung D.-J.,
RA Kwak E., Barlev N.A., Berger S.L., Chow V.T., Roeder R.G.,
RA Azorsa D.O., Meltzer P.S., Suh P.-G., Song E.J., Lee K.-J., Lee Y.C.,
RA Lee J.W.;
RT "Activating signal cointegrator 2 belongs to a novel steady-state
RT complex that contains a subset of trithorax group proteins.";
RL Mol. Cell. Biol. 23:140-149(2003).
CC -!- FUNCTION: May be involved in transcriptional regulation.
CC -!- SUBUNIT: Belongs to the ASC-2/NCOA6 complex (ASCOM), which
CC contains ASC-2/NCOA6, the retinoblastoma-binding protein RBQ-3/
CC RBBP5, alpha- and beta-tubulins, the trithorax group proteins
CC MLL2 and MLL3, and ASH2/ASCL2.
CC -!- SUBCELLULAR LOCATION: Nuclear (Probable).
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=3;
CC Name=1;
CC IsoId=O14686-1; Sequence=Displayed;
CC Name=2;
CC IsoId=O14686-2; Sequence=VSP_008563, VSP_008559;
CC Name=3;
CC IsoId=O14686-3; Sequence=VSP_008560;
CC -!- TISSUE SPECIFICITY: Expressed in most adult tissues, including a
CC variety of hematopoietic cells, with the exception of the liver.
CC -!- MISCELLANEOUS: This gene mapped to a chromosomal region involved
CC in duplications and translocations associated with cancer.
CC -!- SIMILARITY: Belongs to the transcription factor trithorax family.
CC -!- SIMILARITY: Contains 5 PHD-type zinc fingers.
CC -!- SIMILARITY: Contains 1 post-SET domain.
CC -!- SIMILARITY: Contains 1 RING-type zinc finger.
CC -!- SIMILARITY: Contains 1 SET domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF010403; AAC51734.1; -;
DR EMBL; AF010404; AAC51735.1; -;
DR PIR; T03454; T03454.
DR PIR; T03455; T03455.
DR Genew; HGNC:7133; MLL2.
DR MIM; 602113; -;
DR GO; GO:0005634; C:nucleus; TAS.
DR GO; GO:0003700; F:transcription factor activity; TAS.
DR GO; GO:0007048; P:oncogenesis; TAS.
DR GO; GO:0006366; P:transcription from Pol II promoter; TAS.
DR InterPro; IPR003889; FYrich_C.
DR InterPro; IPR003888; FYrich_N.
DR InterPro; IPR000910; HMG_12_box.
DR InterPro; IPR003616; PostSET.
DR InterPro; IPR006118; Recombinase.
DR InterPro; IPR001214; SET.
DR InterPro; IPR001965; Znf_PHD.
DR InterPro; IPR001841; Znf_ring.
DR Pfam; PF00628; PHD; 5.
DR Pfam; PF00856; SET; 1.
DR SMART; SM00542; FYRC; 1.
DR SMART; SM00541; FYRN; 1.
DR SMART; SM00398; HMG; 1.

DR SMART; SM00249; PHD; 7.
DR SMART; SM00508; PostSET; 1.
DR SMART; SM00184; RING; 3.
DR SMART; SM00317; SET; 1.
DR PROSITE; PS50868; POST_SET; 1.
DR PROSITE; PS50280; SET; 1.
DR PROSITE; PS01359; ZF_PHD_1; 5.
DR PROSITE; PS50016; ZF_PHD_2; 5.
DR PROSITE; PS50089; ZF_RING_2; 1.
KW Nuclear protein; Transcription regulation; Coiled coil; Zinc-finger;
KW Repeat; Alternative splicing; Polymorphism.
FT ZN_FING 226 276 PHD-TYPE 1.
FT ZN_FING 229 274 RING-TYPE.
FT ZN_FING 273 323 PHD-TYPE 2.
FT ZN_FING 1102 1155 PHD-TYPE 3.
FT ZN_FING 1152 1202 PHD-TYPE 4.
FT ZN_FING 1229 1284 PHD-TYPE 5.
FT DOMAIN 5121 5242 SET.
FT DOMAIN 5246 5262 POST-SET.
FT DOMAIN 2397 2436 COILED COIL (POTENTIAL).
FT DOMAIN 2788 2809 COILED COIL (POTENTIAL).
FT DOMAIN 2974 3001 COILED COIL (POTENTIAL).
FT DOMAIN 3286 3342 COILED COIL (POTENTIAL).
FT DOMAIN 3437 3476 COILED COIL (POTENTIAL).
FT DOMAIN 3621 3701 COILED COIL (POTENTIAL).
FT DOMAIN 4265 4287 COILED COIL (POTENTIAL).
FT DOMAIN 439 668 15 X 5 AA REPEATS OF S/P-P-P-E/P-P-E/A.
FT REPEAT 442 446 1.
FT REPEAT 460 464 2.
FT REPEAT 469 473 3.
FT REPEAT 496 500 4.
FT REPEAT 504 508 5.
FT REPEAT 521 525 6.
FT REPEAT 555 559 7.
FT REPEAT 564 568 8.
FT REPEAT 573 577 9.
FT REPEAT 582 586 10.
FT REPEAT 609 613 11.
FT REPEAT 618 622 12.
FT REPEAT 627 631 13.
FT REPEAT 645 649 14.
FT REPEAT 663 667 15.
FT DOMAIN 229 326 CYS-RICH.
FT DOMAIN 374 922 PRO-RICH.
FT DOMAIN 1015 1053 ARG-RICH.
FT DOMAIN 1122 1235 CYS-RICH.
FT DOMAIN 1832 2351 PRO-RICH.
FT DOMAIN 2536 2547 GLN-RICH.
FT DOMAIN 2587 2703 PRO-RICH.
FT DOMAIN 2986 4000 GLN-RICH.
FT DOMAIN 3966 4085 PRO-RICH.
FT DOMAIN 4634 4702 PRO-RICH.
FT VARSPLIC 1 305 Missing (in isoform 2).
FT 306 /FTid=VSP_008563.
FT 672 PMEELPAHSWKCKACRVCRCACGAGSALNPSENFENYSLC
FT HRCXKAGGQTIRSVAEQHTFVCSRFSPPEPGDTPTDEPDA
FT LYVACQGPKGHVTSMQKEGPIQCEAKPLGKAGVQLEP
FT QLEAPLNEEMPLLPPPEESPLSPPEESPTSPPEASRLSP
FT PPEELPASPLPEALHLSRPLEESPLSPPEESPLSPPESS
FT PPSLEESPLSPPEESPPSPALETPLSPPEASPLSPPEE
FT SPLSPPEELPTSPPEASRLSPPEESPMSPPEESPMSP
FT PPASRLFPFEEESPLSPPEESPLSPPEASRLSPPEES
FT PMSPPPEESPMSPPEESRLSPLVVSRLSPPEESPLS
FT -> NSPPPEESPMSPPEESRLSPPEESPLSPPEESPLS
FT PPPEASRLSPPEESPMSPPEESPMSPPEESRLSPLPVV
FT SRSPPEESPLSPPEESPTSPPEASRLSPPEESPTSP
FT PPEDSPASPPEDSLMSLPLEESPLLPPEEQLCPRSEGP
FT HLSRPEEPHLSRPEEPHLSQAEPEHLSQPEEPCLCAV
FT PEEHLSQAEQPHLSQPEELHLSQTEEPHLSVPPEEPC
FT LSPQPEESHLSQSEEPCLSPRPEESHLSPELEKPLSPRP
FT EKPPEEPGQCQPAPEELPLFPPEEPFSLPLGEPALSEPGE
FT PPLSPLPEELPLSPSGEPLSQMLMPPDPLPPPLSPIITAA
FT A (in isoform 2).

```
FT VARSPLIC 1454 1454 /FTid=VSP 008559.
FT E -> EGET (in isoform 3).
FT /FTid=VSP 008560.
FT R -> H (in dbSNP:37823356).
FT /FTid=VAR 017115.
SQ SEQUENCE 5262 AA; 564171 MW; 26B7C74CAD417E44 CRC64;

Query Match
Best Local Similarity 22.1%; Score 218.5; DB 1; Length 5262;
Matches 146; Conservative 58; Mismatches 229; Indels 229; Gaps 32;

QY 3 PRNHP-----SRPPSTSLPQGSAP-----LPSAPISSMFM-PQYTMQP 39
Db |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::|
2087 PPSHPDIFRPGSYTDPYAQPPLTPRPQPPPPESCCALPPRSLPSDPFSRVPVSPQSQSS 2146
QY 40 QYPVSPQHTL-----PPLQPHHSQSPAPHSYMGPYRDPDLNRYPASSHDVYASSAAP 92
Db |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::|
2147 QSPLT-PRPLSAEAFCSPTPRF-QSPDPYS---RPPSRP-----QSRDPPFAPLHKP 2194
QY 93 IMP-----HTTVG-----SLP-----PTSFLSHPN----- 112
Db |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::|
2195 PRQPPEVAFKAGSLAHTSLGAGGFPAALPAGPAGELHAKVPSPQPPNFVRSPCTGAFVG 2254
QY 113 -----PQAQAQSQSPHYPPPHSVLPPASSAQSYQPIAPAPPRDRRADFN----- 158
Db |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::|
2255 TPSPMRFTFPQAVGEPSLKP--PVPQPGLPHPHGHNSH---FGPGTLGKQSTNYTVAT 2309
QY 159 -----NGLPSGAFSYDGKPGQWDVPAANGAAPYPGKDSPTQVV-GSQGRGILPSVPG 212
Db |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::|
2310 GNHPGSGPLGPSSGSGTGESYGLSLPRPPSVLPPPADGSLPYLSHGASQSGITSPVEK 2369
QY 213 RATPVTNGVNGTGKNTTIPAKDADG-KFP-CPNCKNTYLHAKHLKRHLRHTGDRPYMCV 270
Db |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::|
2370 REDP-----GTGMGSSLATAELPGTQDPGMSGLSQTLEKQQRQL----- 2411
QY 271 LCKDTFSRSDILKRHFQKCSIRRGNPT-----GATHLSHPNAHVKRSQQQA 316
Db |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::|
2412 -----RELLIRQQIQRNLTLRQEKETAAAAAGAVGPPGSWGAEPSSPAPEQLSRGQ--- 2461
QY 317 AANPVKPVQDEVSTVPENGIPGTTYGEGAVNGNL-----APARPGYAD---HQTMG- 367
Db |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::|
2462 --TPFAGTQDKSSLVGLPFSKLSGPIPGFSFPDDRLSRPPPPATPSSMDVNSRQLVGG 2519
QY 368 -----FPMSSVNGMGR---GQP- 381
Db |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::|
2520 SQAFYQRAYPGSLPLQOQQQQLWQQQQAATAATSMRFAMSARFPSTPGPELGRQALGSPL 2579
QY 382 ---EDAFPGRPHQGAPEWPAQKQSPYL---VQPGADPSGHQLNIDRNIEQVK-QPVVQD 434
Db |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::|
2580 AGISTR LPG--PGEVPVPGAGPAQFIELRHNQKGLGPGGTFFPGQGPQRPFRYPVSED 2637
QY 435 PKRPVMPGHPGHPGELDWTSMFQPAPEGYMFQSMGPGQGEPIHAHVETERKYPTTTAG 494
Db |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::|
2638 PHRLAPEGLRGLA-----VSGLPPQKP-----SAPPAPELNNSLHPTPHTKGPTLPTG 2685
QY 495 QE 496
Db 2686 LE 2687
```

```
RESULT 13
DRPL_RAT
ID DRPL_RAT STANDARD; PRT; 1183 AA.
AC P54258;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Atrophin-1 (Dentatorubral-pallidoluysian atrophy protein).
GN DRPLA.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
```

```
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Cerebellum, and Striatum;
RX MEDLINE=97317138; PubMed=9173996;
RA Loev S.J., Margolis R.L., Young W.S., Li S.-H., Schilling G.,
RA Ashworth R.G., Ross C.A.;
RT "Cloning and expression of the rat atrophin-I (DRPLA disease gene)
RT homologue.";
RL Neurobiol. Dis. 2:129-138(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain, Cerebellum, Hippocampus, and Substantia nigra;
RX MEDLINE=96081227; PubMed=8541849;
RA Schmitt I., Epplen J.T., Riess O.;
RT "Predominant neuronal expression of the gene responsible for
RT dentatorubral-pallidoluysian atrophy (DRPLA) in rat.";
RL Hum. Mol. Genet. 4:1619-1624(1995).
CC -!- SUBUNIT: Interacts with WWP1 and WWP2 (By similarity).
CC -!- TISSUE SPECIFICITY: Predominant neuronal expression, although
CC markedly reduced amounts are found in most other tissues.
CC -!- DEVELOPMENTAL STAGE: Similar expression at all development stages
CC (14.5 dpc, 17.5 dpc, newborns and adults).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U31777; AAA80337.1; -.
DR EMBL; X89453; CAA61623.1; -.
DR InterPro; IPR002951; Atrophin.
DR Pfam; PF03154; Atrophin-1; 2.
DR PRINTS; PR01222; ATROPHIN.
FT DOMAIN 165 171 POLY-PRO.
FT DOMAIN 303 306 POLY-PRO.
FT DOMAIN 377 383 POLY-SER.
FT DOMAIN 387 391 POLY-SER.
FT DOMAIN 440 446 POLY-PRO.
FT DOMAIN 477 480 POLY-HIS.
FT DOMAIN 481 489 POLY-GLN.
FT DOMAIN 502 505 POLY-PRO.
FT DOMAIN 562 572 POLY-SER.
FT DOMAIN 702 705 POLY-PRO.
FT CONFLICT 455 455 N -> S (IN REF. 2).
FT CONFLICT 594 594 F -> L (IN REF. 2).
FT CONFLICT 689 689 P -> R (IN REF. 2).
FT CONFLICT 717 717 T -> M (IN REF. 2).
FT CONFLICT 737 737 A -> V (IN REF. 2).
FT CONFLICT 965 965 MISSING (IN REF. 2).
SQ SEQUENCE 1183 AA; 124778 MW; 7FB9928DCADF9B1F CRC64;

Query Match
Best Local Similarity 22.4%; Score 212.5; DB 1; Length 1183;
Matches 139; Conservative 44; Mismatches 221; Indels 217; Gaps 30;
```

```
QY 7 PSRPPSTSLPQ----GSAPLPSAPISSMFMFPQYTMQPYVSQ-----PHTLPPLPQPHS 57
Db |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::|
166 PSPPPPDSIPRQPESGFEPHPSVPTGYHAP---MEP--PTSRLFQGPFPGAPPPHPOLY 220
QY 58 QSPAPHSYMGQP-----YRDLNRYPASSHDVYASSAAPIM-PH 96
Db |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::|
221 PGSAGGVLGPPMGPKGAAASSVPPSGKGKQHPPTTPIPISSSG--ASGAPPAPKPPN 278
QY 97 TTVG-----SLPPTSFLSHPNP-----QAAQ-----AQQSP-HYPPPHSV-- 131
Db |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::|
279 TPVGAGNLPSAPPATFPHVTNLPPLPPALPLNNASASPPGMGAQPIGHLPSPHAMGQ 338
QY 132 -----LPPASSAQSYQPIAPAPPRDRRADFNGLPSGAFSYSD 170
Db |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::|
339 GMSGLPPGPEKGTLPAPSPHPLPPASS-----APGPP--MRYPYSSCSSSSVAASSS 389
```


QY 171 GKFQGWDPVAANGAAPYPGKDSPTQV-VGSQRRRGILPSVPGRATPVTVNGVNGTGKNTT 229
Db 390 SSAATSYQYPASQTLPSYPHSPFPPTSMVSQNPCKYQPSLPSQA-----VMSQGP 440
QY 230 IPAKDADGKFPKPCNCKNTYLHAKHLKRHLRHTGDRPYMCVLCCKDTFSRSDILKRHFQKC 289
Db 441 PPPPPYGRJ-LPNN-----THPGFP----- 462
QY 290 SRRGNPTGATHLSHPNA---HVKRSQQAAANPVKPVQDEVSVSTVPPNGIPGTTYGEG 346
Db 463 -----PTGGQSTAHPPAPAHHHHQQQQPQP-PQOQHGHGNSGPPPG-----AYPHP 511
QY 347 AVNQN-----GLAPAPG-----YADHQTMGFPMSVNGMGRGQ 380
Db 512 LESSNSHHAHPNMSPSLSRLRPPGPAHLPPSHGQVSYSQAGPNGPPVSSSSSSSGSS 571
QY 381 PEDAFPGGRPH-----QGAPWPQAPKQSPYLVQPGADPSGHQLNIDRNIEQVKQPVVQDP 435
Db 572 SQAAYSCSHPSSSQPGQASYPFP-----VPPITTSSA-----TLSTVIATVASSP 618
QY 436 KRPVMPGHGHPGELDWTSFMFQPAPEGYMFSSQMPGGQEPHAHVETERKYPTTTAGQ 495
Db 619 AGYKTASPPGPP-----QYSKRAPSPGSYKATATPPGYKFGSP-----PSFRGTGT 662
QY 496 ESGMNGLYLASTMSGDGTQVP 516
Db 663 PPGYRG---TSPAPAGGTGTP 680

RESULT 14

KLF2_MOUSE
ID KLF2_MOUSE STANDARD; PRT; 354 AA.
AC Q60843;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Kruppel-like factor 2 (Lung kruppel-like factor).
GN KLF2 OR LKLF.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6 X CBA; TISSUE=Lung;
RX MEDLINE=96025976; PubMed=7565748;
RA Anderson K.P., Kern C.B., Crable S.C., Lingrel J.B.;
RT "Isolation of a gene encoding a functional zinc finger protein
RT homologous to erythroid Kruppel-like factor: identification of a new
RT multigene family."
RL Mol. Cell. Biol. 15:5957-5965(1995).
CC -!- FUNCTION: Binds to the CACCC box in the beta-globin gene promoter
CC and activates transcription.
CC -!- SUBUNIT: Interacts with WWP1 (By similarity).
CC -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC -!- TISSUE SPECIFICITY: PREDOMINANT EXPRESSION IN THE LUNGS AND
CC SPLEEN.
CC -!- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-
CC FINGER PROTEINS.
CC -!- SIMILARITY: Contains 3 C2H2-type zinc fingers.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U25096; AAA86728.1; -.
DR HSSP; P08047; 1SP2.
DR TRANSFAC; T01677; -.

DR MGD; MGI:1342772; Klf2.
DR GO; GO:000515; F:protein binding; IPI.
DR GO; GO:0030528; F:transcription regulator activity; IDA.
DR GO; GO:0045941; P:positive regulation of transcription; IDA.
DR InterPro; IPR007087; Znf_C2H2.
DR Pfam; PF00096; zf-C2H2; 3.
DR ProDom; PD000003; Znf_C2H2; 1.
DR SMART; SM00355; Znf_C2H2; 3.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 3.
DR PROSITE; PS0157; ZINC_FINGER_C2H2_2; 3.
KW Transcription regulation; Activator; Zinc-finger; Metal-binding;
KW DNA-binding; Nuclear protein; Repeat.
FT DOMAIN 60 70 POLY-PRO.
FT DOMAIN 165 169 POLY-PRO.
FT DOMAIN 224 230 POLY-ALA.
FT ZN_FING 271 295 C2H2-TYPE 1.
FT ZN_FING 301 325 C2H2-TYPE 2.
FT ZN_FING 331 353 C2H2-TYPE 3.
SQ SEQUENCE 354 AA; 37700 MW; C4A99D018AC5BAF7 CRC64;
Query Match 7.4%; Score 211.5; DB 1; Length 354;
Best Local Similarity 26.4%; Pred. No. 0.00028;
Matches 92; Conservative 27; Mismatches 133; Indels 97; Gaps 17;
QY 5 NHPSRPPSTSLPQGSAPLPSAPISMPMPQYTMQPVVSQPHTLPLQLPHHSQSAPHS 64
Db 59 NPPEPPPPPPPAFYYPEGAP-----PPYSI-PADSLGTELLRPDLDP--PQGPALHG 109
QY 65 -YMGQPPYRPDLNRYPASSHDVYASSAAPIMPHHTTVGSLPPTFSLSHNPQAAQAQSQSP 123
Db 110 RELLAPPGR--LVKAEPPEVDGGYGCAPLAH-----GPRGLKLEGAPGATGACMRGP 161
QY 124 --HYPPPHSVLP-----PASSAQS-YPOPIAPAPRRDRADFNGLPSGAFSYSDG 171
Db 162 AGRPPPPPDTPPLSPDGPURIPASGPRNPPFPFGPGP-----SFGGPGPALHYGPP 213
QY 172 KPQGW-----PVAANGAAPPYPGKDSPTQVVGSGQRRGILPSVPGRATPVTVNGVNGT 224
Db 214 AFGAFGLFEDAAAAAALGLAP-----PATRGLLTTPSSPLE-- 250
QY 225 GKNTTIPAKDADGKFP-----NCNKTYLHAKHLKRHLRHTGDRPYMCVL- 271
Db 251 -----LLEAKPKRGRRSWPRKRAATHTCSTYNTCGKTYTKSSHLKAHLRTHTGKPYHCNWE 306
QY 272 -CKDTFSRSDILKRHFQKCSIRRGNTGTATHL-----SHPNAHVKR 311
Db 307 GCGWKFARSDELTRHYRK---HTGHRPFQCHLCDRAFSRSDHLALHMKR 352

RESULT 15

PRD1_HUMAN
ID PRD1_HUMAN STANDARD; PRT; 789 AA.
AC O75626;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE PR-domain zinc finger protein 1 (Beta-interferon gene positive-
DE regulatory domain I binding factor) (BLIMP-1) (Positive regulatory
DE domain I-binding factor 1) (PRDI-binding factor-1) (PRDI-BF1).
GN PRDM1 OR BLIMP1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91224495; PubMed=1851123;
RA Keller A.D., Maniatis T.;
RT "Identification and characterization of a novel repressor of beta-
RT interferon gene expression."
RL Genes Dev. 5:868-879(1991).
CC -!- FUNCTION: Transcriptional repressor that binds specifically to the
CC PRDI element in the promoter of the beta-interferon gene. Drives

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 1, 2004, 17:40:49 ; Search time 45 Seconds
(without alignments)
3638.977 Million cell updates/sec

Title: US-10-029-180-8
Perfect score: 2866
Sequence: 1 MDPNRHPSRPPSTSLPQGS.....NGLYLASTMSGDGTQVPARQ 519

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

- Database : SPTREMBL 25.*
- 1: sp_archaea.*
 - 2: sp_bacteria.*
 - 3: sp_fungi.*
 - 4: sp_human.*
 - 5: sp_invertebrate.*
 - 6: sp_mammal.*
 - 7: sp_mhc.*
 - 8: sp_organelle.*
 - 9: sp_phage.*
 - 10: sp_plant.*
 - 11: sp_rodent.*
 - 12: sp_virus.*
 - 13: sp_vertebrate.*
 - 14: sp_unclassified.*
 - 15: sp_rvirus.*
 - 16: sp_bacteriap.*
 - 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	2866	100.0	519	3	Q8J177	Q8J177 emerice
2	274.5	9.6	1724	5	P91019	P91019 caenorhabdi
3	252.5	8.8	722	4	Q8IYCI	Q8IYCI homo sapien
4	240	8.4	723	4	Q8NBB9	Q8NBB9 homo sapien
5	240	8.4	1766	4	Q8NF45	Q8NF45 homo sapien
6	238.5	8.3	1191	4	Q86V38	Q86V38 homo sapien
7	237	8.3	1193	5	Q9VQ94	Q9VQ94 drosophila
8	233	8.1	722	3	Q9C170	Q9C170 neurospora
9	232.5	8.1	3084	12	Q8UZ11	Q8UZ11 pseudorabie
10	231.5	8.1	726	11	Q924S6	Q924S6 mus musculu
11	230.5	8.0	596	11	Q9CVF3	Q9CVF3 mus musculu
12	229.5	8.0	926	5	Q9W3G1	Q9W3G1 drosophila
13	224	7.8	839	16	Q9RX57	Q9RX57 deinococcus
14	222.5	7.8	1175	11	P70200	P70200 mus musculu
15	221.5	7.7	413	4	Q7Z718	Q7Z718 homo sapien
16	221	7.7	412	11	Q9QYG4	Q9QYG4 rattus norv

17	219.5	7.7	695	3	Q8NKC8	Q8nkd8 giberella
18	216.5	7.6	1778	4	Q8IWR5	Q8iwr5 homo sapien
19	216.5	7.6	1858	4	Q8NFF8	Q8nff8 homo sapien
20	216.5	7.6	1858	4	Q8IZD2	Q8izd2 homo sapien
21	216	7.5	389	4	Q86VQ7	Q86vq7 homo sapien
22	216	7.5	398	13	Q802P5	Q802p5 falco colum
23	215	7.5	428	11	Q811U1	Q811u1 rattus norv
24	214.5	7.5	1175	11	Q35126	Q35126 mus musculu
25	213	7.4	389	4	Q8N805	Q8n805 homo sapien
26	213	7.4	1966	5	Q9NHX6	Q9nhx6 drosophila
27	212.5	7.4	2133	5	Q8IQT4	Q8iqt4 drosophila
28	212	7.4	382	4	Q00599	Q00599 homo sapien
29	211.5	7.4	697	3	Q8NK75	Q8nk75 glomerella
30	211	7.4	417	13	Q90XX6	Q90xx6 oncorhynchu
31	211	7.4	691	4	Q86WM7	Q86wm7 homo sapien
32	211	7.4	728	4	Q7Z6T5	Q7z6t5 homo sapien
33	211	7.4	997	5	Q9GYL4	Q9gy14 caenorhabdi
34	211	7.4	1153	5	Q8MQW5	Q8mqw5 drosophila
35	211	7.4	1966	5	Q8IOA6	Q8iga6 drosophila
36	211	7.4	1985	5	Q9VSK5	Q9vsk5 drosophila
37	211	7.4	1985	5	Q8T9N4	Q8t9n4 drosophila
38	211	7.4	1988	5	Q86BH2	Q86bh2 drosophila
39	210.5	7.3	745	16	Q89X06	Q89x06 bradyrhizob
40	210	7.3	389	13	Q802J6	Q802j6 deconychura
41	210	7.3	1350	13	Q91929	Q91929 xenopus lae
42	209.5	7.3	903	16	Q82HF3	Q82hf3 streptomyce
43	209	7.3	378	13	Q802I3	Q802i3 vermivora v
44	208.5	7.3	533	11	Q8CAT6	Q8cat6 mus musculu
45	208.5	7.3	838	5	Q8T5L5	Q8t5l5 anopheles g

ALIGNMENTS

RESULT 1

Q8J177	PRELIMINARY;	PRT;	519 AA.
ID	Q8J177		
AC	Q8J177;		
DT	01-MAR-2003 (Tremblrel. 23, Created)		
DT	01-MAR-2003 (Tremblrel. 23, Last sequence update)		
DT	01-OCT-2003 (Tremblrel. 25, Last annotation update)		
DE	Rfec.		
GN	RFEC.		
OS	Emericella nidulans (Aspergillus nidulans).		
OC	Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;		
OC	Eurotiales; Trichocomaceae; Emericella.		
OX	NCBI_TaxID=162425;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RA	Holtzman D.A.;		
RT	"Enhanced Production of Secondary Metabolites in Filamentous Fungi by		
RT	Ectopic Expression of Regulatory Proteins.";		
RL	Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.		
DR	EMBL; AF467711; AAQ14631.1; -.		
DR	InterPro; IPR007087; Znf C2H2.		
DR	Pfam; PF00096; zf-C2H2; 2.		
DR	SMART; SM00355; Znf C2H2; 2.		
DR	PROSITE; PS00028; ZINC_FINGER_C2H2_1; 1.		
DR	PROSITE; PS0157; ZINC_FINGER_C2H2_2; 1.		
SQ	SEQUENCE 519 AA; 55648 MW; 4AACD0892BA3FA40 CRC64;		

Query Match 100.0%; Score 2866; DB 3; Length 519;
Best Local Similarity 100.0%; Pred. No. 3.4e-196;
Matches 519; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MDPRNHPSRPPSTSLPQGSAPLPSAPISMPMPQYTMQPYVSPHPLPPLPQPHHSQSP 60
Db	1	MDPRNHPSRPPSTSLPQGSAPLPSAPISMPMPQYTMQPYVSPHPLPPLPQPHHSQSP 60
QY	61	APHSYMGQPPYRPDLNRYPASSHDVYASSAAPIMPHTTVGSLPPTSLSHPNPQAQAAQ 120
Db	61	APHSYMGQPPYRPDLNRYPASSHDVYASSAAPIMPHTTVGSLPPTSLSHPNPQAQAAQ 120

QY 121 QSPHYPPHSHSVLPASSAQSYQPIAPAPRRDRRADFNGLPSGAFSYSDGKPKQGWDPVA 180
Db 121 QSPHYPPHSHSVLPASSAQSYQPIAPAPRRDRRADFNGLPSGAFSYSDGKPKQGWDPVA 180
QY 181 ANGAAPYPGKDSPTQVVGSGRRGILPSVPGRATPVTVNGVNGTGKNTTIPAKDADGKFP 240
Db 181 ANGAAPYPGKDSPTQVVGSGRRGILPSVPGRATPVTVNGVNGTGKNTTIPAKDADGKFP 240
QY 241 CPNCNKTYLHAKHLKRHLRHTGDRPYMCVLCCKDTFSRSDILKRHFQKCSIRRGNTGAT 300
Db 241 CPNCNKTYLHAKHLKRHLRHTGDRPYMCVLCCKDTFSRSDILKRHFQKCSIRRGNTGAT 300
QY 301 HLSHPNAHVKRSQQQAAANPVQDEVSTVPPNGIPGTTYGEGAVNGNGLAPARPGY 360
Db 301 HLSHPNAHVKRSQQQAAANPVQDEVSTVPPNGIPGTTYGEGAVNGNGLAPARPGY 360
QY 361 ADHQTMGFPMSVNGMGRGQPEDAFPGGRPHQGAPWPQAPKQSPYLVQPGADPSGHQLNI 420
Db 361 ADHQTMGFPMSVNGMGRGQPEDAFPGGRPHQGAPWPQAPKQSPYLVQPGADPSGHQLNI 420
QY 421 DRNIEQVKQPVVQDPKRPVMPGHPGHPGELDWTSMFQPAPEGYMFSSQMPGGQEPPIAH 480
Db 421 DRNIEQVKQPVVQDPKRPVMPGHPGHPGELDWTSMFQPAPEGYMFSSQMPGGQEPPIAH 480
QY 481 VETERKYYPPTTAGQESGMNGLYLASTMSGDGTVPARQ 519
Db 481 VETERKYYPPTTAGQESGMNGLYLASTMSGDGTVPARQ 519

RESULT 2
P91019 ID P91019 PRELIMINARY; PRT; 1724 AA.
AC P91019;
DT 01-MAY-1997 (TrEMBLrel. 03, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein.
GN C01G8.9.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RX MEDLINE=99069613; PubMed=9851916;
RA Waterston R.;
RT "Genome sequence of the nematode C. elegans: a platform for
investigating biology. The C. elegans Sequencing Consortium.";
RL Science 282:2012-2018(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Du Z., Gattung S.;
RT "The sequence of C. elegans cosmid C01G8.";
RL Submitted (JAN-1997) to the EMBL/GenBank/DDBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Waterston R.;
RL Submitted (APR-2002) to the EMBL/GenBank/DDBJ databases.
DR EMBL; U80439; AAB37646.3; -.
DR PIR; T29266; T29266.
DR WormPep; C01G8.9a; CE30415.
DR GO; GO:0005622; C:intracellular; IEA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR InterPro; IPR001606; ARID.
DR InterPro; IPR006031; XYPX.
DR Pfam; PF01388; ARID; 1.
DR Pfam; PF02162; XYPX; 7.
DR SMART; SM00501; BRIGHT; 1.
KW Hypothetical protein.
SQ SEQUENCE 1724 AA; 184940 MW; A981E4CEE388EFB8 CRC64;

Query Match 9.6%; Score 274.5; DB 5; Length 1724;
Best Local Similarity 22.6%; Pred. No. 7.1e-11;
Matches 156; Conservative 53; Mismatches 188; Indels 293; Gaps 36;
QY 16 PQGSAPLPSAPISSMPMPQYTMQ-----PQYVVSQP-----HTLPPLQ-----PH 55
Db 24 PKKDDQVPPADASMTPTQQCCQQLPPPPQQPYEHPAQHPPQHHSVPFNSFAPPGQH 83
QY 56 HSQSPA-----PHSY-MGQPPYRDLNRYPASSHDVYA----- 87
Db 84 HPQHPGMPMEWRPPGAEYQMPPGYPAGYPPYGMPPRHHGYPHPAYGYPPPGAPYGYPP 143
QY 88 -SSAAPIMPHTTVGSLPPTSFLSHPNP---QAQAQAQSPHYPPH-----SVLPASSA 138
Db 144 QMMRPPMMAAPGDMVRMPP-----GPTPTWAAQQAASRSVPPKEGPNGNPATPSSS 198
QY 139 QSYQPPIAPAPRRDRRADFNGLPSGAFSYSDGKPKQGWDP----- 178
Db 199 QPSPSPASSIAEESLDD---KPSGTKMPAQPPPPQHPPPPPQIMSPMPQAPPSQ 254
QY 179 -VAANGAAPYPGKDSPT-----QVVG-----SQRRG----- 205
Db 255 ATSSSAASVAAPDTPKVSQKVLKAELEKLVGPTLHNPVKMAERRGFFERLIEFCE 314
QY 206 -----ILPSVPGRATFVTN---GVNGTGKNTTIPAKDADGKPPCPNCNK----- 247
Db 315 HNGEPLTMVPQVSKQSIDLHLRLYIGVRAGGFGQV-TKDKYKWLCTEANPDLAESSAAG 373
QY 248 YLHAKHLKRHL-----RETGDRPYMCVLCCKDTFSRSDILKRHFQKCSIRRGNTGATHLS 303
Db 374 YQLRKHYQRHLLMLECRETGRNP-----EDEVAFADKMKRQ-----RKREPAG----- 417
QY 304 HPNAHVKRSQQQAAANPVQVQ-DEVSTVPPPPNGIPGTTYGEGAVNGNGLAPARPGYA 361
Db 418 --NA-----AAAAAQAQGDQKQ-----HGAPG-----GSGAPPPGP----- 450
QY 362 DHQTMGFPMSVNGMGRGQPEDAFPGGRPHQGAPW--PQAPKQSPYLVQPGADPSGHQLN 419
Db 451 -----PGAYPGNGAPGGPGPPGPPQFP--GHPGMDPNVHYQ 484
QY 420 IDRNIEQVKQPVVQDPKRPVMPGHPGHPGELDWTSMFQVQ-----APEGYMFSSQMPGG- 473
Db 485 -----QHGMMPHPGHPG-----YPPQHMNAFSPGQYPGHQRPFGGP 521
QY 474 -----QEPIHAHV-----ETERKYY----- 488
Db 522 GGPPPGPQAMRAPMPQHMQEMDEHQRYAHQQAHAHHAQQAQAQAQAQAQAQAQAQASTP 581
QY 489 -PTTTAGQESGMNGLYLASTMSGDGTVPQPA 517
Db 582 APSTTPAPQAASSS--QAAPTSGANT-QPA 608

RESULT 3
Q8IYCI ID Q8IYCI PRELIMINARY; PRT; 722 AA.
AC Q8IYCI;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Similar to zinc finger protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RA Strausberg R.;
RL Submitted (JUL-2002) to the EMBL/GenBank/DDBJ databases.
DR EMBL; BC036105; AAH36105.1; -.
DR GO; GO:0016021; C:integral to membrane; IEA.

DR GO; GO:0004969; F:histamine receptor activity; IEA.
DR GO; GO:0007186; P:G-protein coupled receptor protein signalin. . .; IEA.
DR InterPro; IPR003980; H3 receptor.
DR InterPro; IPR007087; ZnF_C2H2.
DR Pfam; PF00096; zf-C2H2; 8.
DR PRINTS; PR01471; HISTAMINEH3R.
DR SMART; SM00355; ZnF_C2H2; 9.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 5.
DR PROSITE; PS50157; ZINC_FINGER_C2H2_2; 6.
SQ SEQUENCE 722 AA; 76881 MW; 8CA879DAB9F78759 CRC64;

Query Match 8.8%; Score 252.5; DB 4; Length 722;
Best Local Similarity 24.3%; Pred. No. 9.2e-10;
Matches 146; Conservative 45; Mismatches 152; Indels 257; Gaps 36;

QY 7 PSRPPSTSLQGSAPLPSAPISSMPMPQYTMQPYVSPQHTLPLQPHHSQSP-----A 61
| | | | : : : : | | | | : : : : | | | | : : : : |
Db 215 PERPLAAT---SAAPP-----QPQPQPQPQE-PRSVQPQPEPEPEPEREATPTTAPAA 264
| | | | : : : : | | | | : : : : | | | | : : : : |
QY 62 PHSYMGQPPYRPD-----LNRYPAS-----SH 83
| | | | : : : : | | | | : : : : | | | | : : : : |
Db 265 PE-----EPPAPPEFRQCVCQSFTQSWFLKGHRKHKASFHDHACVPVCGRCFKEPWFLKNH 320
| | | | : : : : | | | | : : : : | | | | : : : : |
QY 84 -DVYASSAAPIMPTTVGSLPPTSFSLHPNP-QAQAQAQSPH-----YPPHSVLPPAS 136
| | | | : : : : | | | | : : : : | | | | : : : : |
Db 321 MKVHASKLGP-----LRAPGASGPAPARAPQPPDLGLLAYEP-----LGPAL 361
| | | | : : : : | | | | : : : : | | | | : : : : |
QY 137 SAQSYPOPIAPAP-PRDRRADFNGLPSGAFSY-----SDGKPGQWDFVAANGAAPPYPGK 190
| | | | : : : : | | | | : : : : | | | | : : : : |
Db 362 L-----LAPAPTAEER-----PPSLLGLSLRAGEGRPNG-----EGAEPPGR 402
| | | | : : : : | | | | : : : : | | | | : : : : |
QY 191 D-----SPRTQVVGSGRR-----GILPSV-----PGR 213
| | | | : : : : | | | | : : : : | | | | : : : : |
Db 403 SFGGFRPLSSALPARARRHRAEEPEEEVEAEETWARGRSLSGLSLASLHPRPGEFGH 462
| | | | : : : : | | | | : : : : | | | | : : : : |
QY 214 -----ATPVTNG--VNGT---GKNTTIPAKDADGKFPDPCNCKNTYLAHAKHLK 255
| | | | : : : : | | | | : : : : | | | | : : : : |
Db 463 SASAAGAARSTATQENGLLVGTRPGEGRGAT--GKD-----CPFCGKSFSAHHLK 514
| | | | : : : : | | | | : : : : | | | | : : : : |
QY 256 RHLLRHTGDRPYMCVLCKDTFSRSDILKRHFQKCSIRRGNTGTATHLSHPNAHVKRS--- 312
| | | | : : : : | | | | : : : : | | | | : : : : |
Db 515 VHLRVHTGERPYKPCPCDYAGTQSGSLKYLQRRHREQRSAGPGPPPEPPPSQSGSAP 574
| | | | : : : : | | | | : : : : | | | | : : : : |
QY 313 QQQAAANPVKQDEVSVSTVPPNGIPGTTYGEGAVNGNGLAPARPGYADHQTMTGPFMSS 372
| | | | : : : : | | | | : : : : | | | | : : : : |
Db 575 QSGAKPSQPATWVEGASSRPPS-----SGAGPSRRKPKASPG---RTLK----- 617
| | | | : : : : | | | | : : : : | | | | : : : : |
QY 373 VNGMGRGQPE-----DAFPGGRPHQ-----GAPW----- 396
| | | | : : : : | | | | : : : : | | | | : : : : |
Db 618 -NGRG-GEAEPLDSLRAAGPGGEAGPGGALHRLCFPFATGAPELMALHLQVHHSRRARG 675
| | | | : : : : | | | | : : : : | | | | : : : : |
QY 397 ---PQAPKQSPYLQPGADPSGHQNLNIDRNIEQVKQPVVQDPKRPVMPGHGPG-----HPGE 449
| | | | : : : : | | | | : : : : | | | | : : : : |
Db 676 RRPQADASPPY---ARVPSG-----ETPPSPSQEGEGSLSRPGE 714
| | | | : : : : | | | | : : : : | | | | : : : : |

RESULT 4
Q8NBB9 PRELIMINARY; PRT; 723 AA.
ID Q8NBB9
AC Q8NBB9;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein FLJ33731.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RA Tashiro H., Yamazaki M., Watanabe K., Kumagai A., Itakura S.,
RA Fukuzumi Y., Fujimori Y., Komiyama M., Sugiyama T., Irie R.,

RA Otsuki T., Sato H., Ota T., Wakamatsu A., Ishii S., Yamamoto J.,
RA Isono Y., Kawai-Hio Y., Saito K., Nishikawa T., Kimura K.,
RA Yamashita H., Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K.,
RA Wagatsuma M., Murakawa K., Kanehori K., Takahashi-Fujii A., Oshima A.,
RA Sugiyama A., Kawakami B., Suzuki Y., Sugano S., Nagahari K.,
RA Masuho Y., Nagai K., Isogai T.;
RT "NEDO human cDNA sequencing project."
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK091050; BAC03574.1; --
DR InterPro; IPR001660; SAM.
DR Pfam; PF00536; SAM; 1.
KW Hypothetical protein.
SQ SEQUENCE 723 AA; 78133 MW; B279155820EA99AA CRC64;

Query Match 8.4%; Score 240; DB 4; Length 723;
Best Local Similarity 22.6%; Pred. No. 7.1e-09;
Matches 132; Conservative 54; Mismatches 198; Indels 200; Gaps 27;

QY 19 SAPLPSAPISSMPMPQYTMQPYVSPQHTLPLQPHHSQSPAPHSYMGQPPYRPDLNRY 78
| | | | : : : : | | | | : : : : | | | | : : : : |
Db 84 SMPPPFVYSQMPPLPTMPP--PVLPPSLPPPVM--PALPATVPPGMPP----- 131
| | | | : : : : | | | | : : : : | | | | : : : : |
QY 79 PASSHDVYASSAAPIMPTTVGSLPPTSFSLHPNPQAQAQAQSPHYPPH-----SV 131
| | | | : : : : | | | | : : : : | | | | : : : : |
Db 132 -----PVMPPSLPTSVPPPGM-----PPSLSSAGPPPVLPPLSFSAGPPV 173
| | | | : : : : | | | | : : : : | | | | : : : : |
QY 132 LPPASSAQSYPOPIAPAPPRDRRADFNGLPSGAFSYSDGKPGQWDFVAANGAAPPYPGK 191
| | | | : : : : | | | | : : : : | | | | : : : : |
Db 174 LPPPSLSSTAPPVMPPLPLS-SATPPPGIP-----PPGVPGQIPPOLT--AAPVPPAS 224
| | | | : : : : | | | | : : : : | | | | : : : : |
QY 192 SPRTQVVGSGRRGILPSVPGRATPVTVNGVNGTGKNTTI-----PAKDADGKFPDPCN 243
| | | | : : : : | | | | : : : : | | | | : : : : |
Db 225 SSQSSQVPEKPRPALP-----TPVS---FGSAPPTTYHPPLQSAGPSEQVNSKAP--- 272
| | | | : : : : | | | | : : : : | | | | : : : : |
QY 244 CNKTYLHAKHLKRLLRHTGDRPYMCVLCKDTFSRSDILKRHFQKCSIRRGNTGTATHLS 303
| | | | : : : : | | | | : : : : | | | | : : : : |
Db 273 -----LSKSAL-----PY-----SSFSSDQGL----- 289
| | | | : : : : | | | | : : : : | | | | : : : : |
QY 304 HPNAHVKSQQQAANPVKPVQD---EVSVTVPPNGIPGTTYGEG----- 346
| | | | : : : : | | | | : : : : | | | | : : : : |
Db 290 -----GESSAAPSQPITAVKDMPVRSGLLPDP--PRSSYLESRGRFDGPRRFED 339
| | | | : : : : | | | | : : : : | | | | : : : : |
QY 347 -----AVNGNGLAPARPGYADHQTMG-----FPMSSVNG 375
| | | | : : : : | | | | : : : : | | | | : : : : |
Db 340 LGSRCGEGRPKGRFEGNEPDGPRPYEGHPAEGTKSKWGMIPRGASQFYITPSTSLSP 399
| | | | : : : : | | | | : : : : | | | | : : : : |
QY 376 MGRGQ---PEDAFPGGRPHQGPWQAPKQSPYLQPGADPSGHQ-----LNID 421
| | | | : : : : | | | | : : : : | | | | : : : : |
Db 400 RQSGPQWKGPKPAF--GQQHQQPKSQAEPLSGN-KEPLADTSSNQCKNFKMQSAAFSIA 456
| | | | : : : : | | | | : : : : | | | | : : : : |
QY 422 RNIEQVK-----QPVVQDPKRPVMPGHGHPHGGELDW-TSMFQPAPEGYMFQSMPG 472
| | | | : : : : | | | | : : : : | | | | : : : : |
Db 457 ADVKDVKAAQSNENLSDSQQEPKSEVSEGPVPSNWDQNVQSMETQIDKAQAVTQPVPL 516
| | | | : : : : | | | | : : : : | | | | : : : : |
QY 473 GQEPHHAHVETERKYYPYTTTAGQESGMNGLYLASTMSGDGTQVP 516
| | | | : : : : | | | | : : : : | | | | : : : : |
Db 517 ANKPVPA-----QSTFPSKTGMEGGT--AVATSSLTADNDFKP 553
| | | | : : : : | | | | : : : : | | | | : : : : |

RESULT 5
Q8NF45 PRELIMINARY; PRT; 1766 AA.
ID Q8NF45
AC Q8NF45;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE FLJ00353 protein (Fragment).
GN FLJ00353.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]

```
RP SEQUENCE FROM N.A.
RC TISSUE=Spleen;
RA Jikuya H., Takano J., Kikuno R., Nagase T., Ohara O.;
RT "The nucleotide sequence of a long cDNA clone isolated from human
RT spleen.";
RL Submitted (JUL-2002) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AK090435; BAC03416.1; -.
FT NON TER 1
SQ SEQUENCE 1766 AA; 200339 MW; 50E4B0150FA73CAF CRC64;

Query Match      8.4%; Score 240; DB 4; Length 1766;
Best Local Similarity 22.6%; Pred. No. 2.1e-08;
Matches 132; Conservative 54; Mismatches 198; Indels 200; Gaps 27;

QY 19 SAPLPSAPISSMPMPQYTMQYQPVVPSQPHTLPLQLPHHSQSPAPHSYMGQPPYRPDLNRY 78
Db 141 SMPPPFVYSQMPPLPTMPP--PVLPPSLPPVMP--PALPATVPPPGMPP-----188

QY 79 PASSHDVYASSAAPIMPHTTTVGSLLPTSFSLSHPNPQAQAQSQSPHYPPPH-----SV 131
Db 189 -----PVMPPSLPTSVPPPGM-----PPSLSSAGPPFVLPPLPSLSSAGPPPV 230

QY 132 LPPASSAQSYQPIAPAPRRDRRDNFNNGLPSGAFSYSDGKQWDPPVAANGAAPPYDGD 191
Db 231 LPPPSLSSTAPPVMPPLPLLS-SATPPPGIP-----PPGVPOGIPPLQLT--AAPVPPAS 281

QY 192 SPRTQVVGSGRRGILPSVPGRATPVTVNGVGTGKNTTI-----PAKDADGKFPCCN 243
Db 282 SSQSSQVEKPRPALP-----TPVS--FGSAPPTTYHPPLQSAQSEQVNSKAP---329

QY 244 CNKTYLHAKHLKRHLRHTGDRPYMCMVLCCKDTFSRSDILKRHFQKCSIRRNPTGATHLS 303
Db 330 -----LSKSAL-----PY-----SSFSSDQGL-----346

QY 304 HPNAHVKRSQQQAANPVKPVQD---EVSSTVPPPNGIPGTTYGEG-----346
Db 347 -----GESSAAPSQPITAVKDMFVRSGLLPDP---PRSSYLESPRGPRFDGPRRFED 396

QY 347 -----AVNGNGLAPARPGVADHQTMG-----FPMSSVNG 375
Db 397 LGSRCGPRPKGRFEGNRPDGRPRYRHEGHPAEGTSKWMIPRGEPASQFYITPSTLSLP 456

QY 376 MGRGQ-----PEDAFPGGRPHQGWAPQAPKQSPYLVPQGDPSGHQ-----LNID 421
Db 457 RQSGPQWKGPAPF--GQHQQQPKSQABFLSGN-KEPLADTSSNQKNFKMQSAAFSIA 513

QY 422 RNIEQVK-----QPVVQDKPRFVMPGHPGHPGELDW-TSMFQQAPEGYMFQSMPG 472
Db 514 ADVKDVKAAQSNENLSDSQEPPKSEVSEGVPEPSNWDQNVQSMETQIDKAQAVTQPVPL 573

QY 473 QGEPIHAHVETERKYPTTTAGQESGMNGLYLASTMSGDGTVP 516
Db 574 ANKFVPA-----QSTFPKTKGMEGGT--AVATSSLTADNDFKP 610
```

RESULT 6

```
Q86V38 PRELIMINARY; PRT; 1191 AA.
AC Q86V38;
DT 01-JUN-2003 (Tremblrel. 24, Created)
DT 01-JUN-2003 (Tremblrel. 24, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE DRPLA protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
```

```
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RA Strausberg R.;
RL Submitted (MAY-2003) to the EMBL/GenBank/DDBJ databases.
DR EMBL; BC051795; AAH51795.1; -.
DR InterPro; IPR002951; Atrophin.
DR Pfam; PF03154; Atrophin-1; 1.
DR PRINTS; PRO1222; ATROPHIN.
SQ SEQUENCE 1191 AA; 125541 MW; 4301148834EA6714 CRC64;
```

Query Match 8.3%; Score 238.5; DB 4; Length 1191;
Best Local Similarity 22.1%; Pred. No. 1.7e-08;
Matches 123; Conservative 51; Mismatches 197; Indels 185; Gaps 23;

```
QY 7 PSRPPSTSLPQGSAPLPSAPISSMPMPQYTMQYQPVVPSQPHTLPLQLPHHSQSPAPHSYM 66
Db 272 PTKPPTT--PVGGNLPSPAP-----PPANFHTVTPNLPP-----303

QY 67 GQPPYRPDLNRYPASSHDVYASSAAPIMPHTTTVGSLLPTSFSLSHPNPQAQA-----119
Db 304 --PPALRPLNNASASPPGL---GAQPLPGHL-----PSPHAMGQGMGLPPG 345

QY 120 -QSSPHY-PPPHSVLPPASSAQSYQPIAPAPRRDRRDNFNNGLPSGAFSYSDGKPGQWD 177
Db 346 PEKGTPLAPSPHS-LPPASS-----APAPPMRFYSSSSSSSSSSSSSSSSSSAS 397

QY 178 PVAANGAAP-YPGKDSPTQV-VGSQRRGILPSVPGRATPVTVNGVGTGKNTTIPAKDA 235
Db 398 PFPASQALPSYPHSFPPFTSLSVSNQPPKYTQPSLPSQAV-----WS 439

QY 236 DGKFPCCNCKTYLHAKHLKRHLRHTGDRPYMCMVLCCKDTFSRSDILKRHFQKCSIRRN 295
Db 440 QGPPPPPPYGRLLANSN-----AHGPFPP-----PSTGAQS-----TAH 473

QY 296 PTGATHLSHPNAHVKRSQQQAANPVKPVQDEVSTVPPPNGIPGTTYGEGAVNGN-----351
Db 474 PFWSTHHHHHQQQQQQQQQQQQQQQQQQQQQHHGNSGPPPPGAPPHPLEGGSSHHAPYAM 533

QY 352 -----GLAPARPG-----YADHQTMGFPMSVNGMGRGQPEDAFPGGRPH---391
Db 534 SPSLGSLRPPYPGPAHLPPPHSQQVSYSQAGPNPVPVSSSSSSSSSSSSSSSSSSPSQ 593

QY 392 --QGAPWFOAP-----KQSPYLVPQADPSGHQLNIDRNIEQVKQPVVQDKRPFVM 440
Db 594 GPQGAPYPPFPVPTVTTSATLSTVIATVASSPAGY-----KTASP 634

QY 441 PGHPGHPGELDWTSMFQPAPEGYMFSSQMPGGQEPPIHAHVETERKYPTTTAGQESGMN 500
Db 635 PGPPPYGKRAPSPGAYKTATPPGY-----KPGSP-----PSFRTGTPPGYR 675

QY 501 GLYLASTMSGDGTVP 516
Db 676 G---TSPPAGPGTFKP 688
```


QY	155	AD-----FNN-----GLPSGAFSYSDGKP-----QGWD	178
Db	494	ADNSPPGLTYSNPAINMARQQHHEMLELSRHGTP---LSTVEGSPALNPMNISMGSYGP	550
QY	179	VAANGAAPYPGKDSPTQVVVGSQRRRGILPSVPGRATPTVNGVGTGKNTTIPAKDADGK	238
Db	551	LNDDELAPMSEMQ-----SQGHPGVI-----RRARSAT--VSELG-----PYPKSHS	592
QY	239	FPCPNCNKTYLHAKHLKRLHRLHRTGDRPYMCVLCKDTFSRSDILKRHFQKCSIRGNPTG	298
Db	593	CPIPCSGRVFKRLEHLKRVHRTHTQERPVCYCSQKAFSRSDNLAQ-----	638
QY	299	ATHLSHPNAHVKRSQQQAAANPVKPVQDEVSSTVPPPNIGIPGTTYGEA-VNGNGLAPAR	357
Db	639	-----HKRTHDRGDGSEGTNSLSEEEEEYS-----PTSEGGYVHASGAPSNNGNSTPQS	689
QY	358	PGYADHQTMGFPMS 371	
Db	690	TMYNLQTLSPMT 703	
RESULT 9			
Q8UZ11	PRELIMINARY; PRT; 3084 AA.		
ID	Q8UZ11		
AC	Q8UZ11;		
DT	01-MAR-2002	(TReMBLrel. 20, Created)	
DT	01-MAR-2002	(TReMBLrel. 20, Last sequence update)	
DT	01-MAR-2003	(TReMBLrel. 23, Last annotation update)	
DE	UL36 protein.		
OS	Pseudorabies virus (strain Kaplan) (PRV).		
OC	Viruses; dsDNA viruses, no RNA stage; Herpesviridae;		
OC	Alphaherpesvirinae; Varicellovirus.		
OX	NCBI_TaxID=33703;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=Kaplan;		
RA	Klupp B.G., Fuchs W., Granzow H., Nixdorf R., Mettenleiter T.C.;		
RT	"The pseudorabies virus UL36 tegument protein physically interacts with the UL37 protein.";		
RL	Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.		
DR	EMBL; AJ422133;	CAD19511.1; --	
DR	InterPro; IPR006928;	Herpes_teg_N.	
DR	InterPro; IPR005210;	Herpes_UL36.	
DR	InterPro; IPR002965;	P-rich extensn.	
DR	Pfam; PF04843;	Herpes_teg_N; 1.	
DR	Pfam; PF03586;	Herpes_UL36; 1.	
DR	PRINTS; PR01217;	PRICHEXTENS.	
SQ	SEQUENCE 3084 AA; 324401 MW; ECAD9E1E3DC22D1A CRC64;		
Query Match			
Best Local Similarity 8.1%; Score 232.5; DB 12; Length 3084;			
Matches 145; Conservative 41; Mismatches 245; Indels 171; Gaps 29;			
QY	6	HPSRPPSTSLPQGSAPLPSAPISSMMPQYTMQPYVVSQPHLPLQLPHHSQSPAPHSY	65
Db	2269	YPATAPAPETAPPEAPQAPPAKATPP---QPQGP-----PPPQPPSAQAPPAQ--	2316
QY	66	MGQPPYRPDLNRYPASSHDVYASSAAPIMPHTTVGSLPPT---SFLSHPNPQAQAQQS	122
Db	2317	--KPPAQP-----ATAAATTAPKATPQTQPTTQATQTATAPPPSAATAAAQV	2361
QY	123	PHYP-----PPHSVLPPASSAQ-SYQPIA-PAPRRDRADFNGLPSGAFSY	168
Db	2362	PPQPPSSQPAKPRGAPPAPPAPPPPSAQTTLPRPAAPPAPP-----PSAQTTL	2411
QY	169	SDGKPGQWDVVAANGAAPYPGKDSPTQVVGSGQRRGILPSVPGRATPTVNGVGTGK--	226
Db	2412	PRPAP---PPPSAPAAATPTPAPGPAPSAKSDGDRIVEPK--AGAPPDVRDAKFGKVA	2466
QY	227	NTTIPAKDAD-----GKFPCCNCNKTYLHAKHLKRLHRLHRTGDRPYMCVLCKDTFSRSD	280
Db	2467	GAASGRPLDDERYKAGR--ADKCEHTVGPDQHRTRSPPGTSDSEP-----ASASSSAAS	2519

DR	EMBL; AK008516; BAB25714.1; -.	
DR	MGD; MGI:1917140; Zfp219.	
DR	InterPro; IPR007087; Znf_C2H2.	
DR	Pfam; PF00096; zf-C2H2; 6.	
DR	SMART; SM00355; Znf_C2H2; 7.	
DR	PROSITE; PS00028; ZINC_FINGER_C2H2_1; 4.	
DR	PROSITE; PS0157; ZINC_FINGER_C2H2_2; 4.	
KW	Metal-binding; Zinc; Zinc-finger.	
FT	NON TER	1
SQ	SEQUENCE	596 AA; 63571 MW; 49C750FDC6760DFE CRC64;
Query Match		
Best Local Similarity		
Matches 128; Conservative 48; Mismatches 202; Indels 165; Gaps 28;		
Qy	16 PQGSAPLPSAPISSMMPQYTMQPYVSPHTLPPLQPHHSQSPAPHSYMGQPPYRPDL	75
Db	92 PLAASTPEPPPPPPQEQPRSALEPE-PEPEPRPEPDREANFAPTPAPE---	EPPAPPEF 147
Qy	76 N-----RYPASSHDVYASSAAPI-----MPHTTVGSLPPTS	106
Db	148 RCQVCGSQFTQSWFLKGHRKHKASFDHACPVCGRCFKEPWLKNHMKVHTS--	KLGP-- 203
Qy	107 FLSHNP-QAQAQAQSPH-----YPP--PHSVLPASSAQSYQP-	-----IAPA 148
Db	204 -LRAFGSAPARAPQPPDLSLLAYEPLGPALLAPAPARERREPSLLGLYSVRAGEV	262
Qy	149 PPRDRADFNGLPSGAFSYSDGKP-----QGWDPVA	180
Db	263 RPNGEGAD-----PGGRSYGGFRPLPSALPNRARRHRTPEPEEEVEAEESWAEGR	317
Qy	181 ANGAA-----PYGKDSR-TQVVGSGRRGILPSVPGRATPVNG--VNGTGKNTTIPAK	233
Db	318 SLGSLTSLHPNPGEGSQPAPAGTQAR-----STATQENGLLVGGTRSEA----	GR 366
Qy	234 DAGKEPCPCNCNTYLHAKHLKRHLRHTGDRPYMCVLCCKDTFSRSDILKRHFQKCSIRR	293
Db	367 GATGK-DCPFCGKFSRSAHHLKVHLRVHTGERPYKCPHCDYAGTQSGSLKYHLQRHREQ	425
Qy	294 GNPTGATHLSHPNAHVKRS--QQQAAANPKVPQDEVSVSTVP--PPNGIPGTTYGEGAVN	349
Db	426 RSSAGPGPPPEPPPPSQRGLQPQSGAKPTQASATWVEGTASTRPPSSSTGP-----	477
Qy	350 GNGLAPARPGYADHQTGMGFFPMSSVNGMRGQPED-----AFPGRP-----	HQ----- 392
Db	478 GSRRKPASPGRTLNRGR-----GEAEPLDSLRAAGPGGEAGAGALHRCLECPFA	528
Qy	393 -GAPWQA-----PKQSPYLVPQGADPSGHQLNIDRNIEQVKQPVVQDPKRPVM--	PG 442
Db	529 TGAPELMALHLQVHSSRRARRGRQRADTSP-----TYVRAPSGETPPSPPLEEEG	579
Qy	443 HPG 445	
Db	580 SPG 582	
RESULT 12		
Q9W3G1	PRELIMINARY; PRT; 926 AA.	
ID	Q9W3G1	
AC	Q9W3G1;	
DT	01-MAY-2000 (TReMBLrel. 13, Created)	
DT	01-MAY-2000 (TReMBLrel. 13, Last sequence update)	
DT	01-OCT-2003 (TReMBLrel. 25, Last annotation update)	
DE	CG10555 protein.	
GN	CG10555.	
OS	Drosophila melanogaster (Fruit fly).	
OC	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;	
OC	Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;	
OC	Ephydroidea; Drosophilidae; Drosophila.	
OX	NCBI_TaxID=7227;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=Berkeley;	

RX	MEDLINE=20196006; PubMed=10731132;	
RA	Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,	
RA	Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,	
RA	George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,	
RA	Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,	
RA	Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,	
RA	Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,	
RA	Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,	
RA	Ballev R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,	
RA	Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,	
RA	Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,	
RA	Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,	
RA	Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,	
RA	de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,	
RA	Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,	
RA	Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,	
RA	Fosler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,	
RA	Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,	
RA	Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,	
RA	Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,	
RA	Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,	
RA	Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,	
RA	Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,	
RA	Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,	
RA	Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,	
RA	Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,	
RA	Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,	
RA	Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,	
RA	Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,	
RA	Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,	
RA	Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,	
RA	Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,	
RA	Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,	
RA	Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,	
RA	Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,	
RA	Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,	
RA	Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;	
RT	"The genome sequence of Drosophila melanogaster.";	
RL	Science 287:2185-2195(2000).	
DR	EMBL; AE003443; AAF46366.1; -.	
DR	FlyBase; FBgn0030034; CG10555.	
DR	InterPro; IPR007726; SSXT.	
DR	Pfam; PF05030; SSXT; 1.	
SQ	SEQUENCE 926 AA; 93004 MW; 6FD5B9F77C36C006 CRC64;	
Query Match		
Best Local Similarity		
Matches 142; Conservative 44; Mismatches 229; Indels 205; Gaps 30;		
Qy	4 RNHP-----SRPPSTSLP-QGSAPLPS-----APISSMMPQYTMQP	39
Db	351 QQHPVYRNAQGQGPAGQVPGQGGVQSVINPNAEPNQRPNNGLSGPNPQQQQQ	410
Qy	40 QYVVSQPHTLPPLQPHHSQSPAPHSYMGQPPYRPLNRYPASSHDVYASSAAPIMPH	99
Db	411 PQFGGQQ---PPNQQQQQQQTGP---GGPGQP-----GAGGPGVP----	445
Qy	100 GSLPPTSLSHPNPQAQAQAQSQSPHYPP-----PHSVLPASSAQSYPOIAPAPPRD	152
Db	446 ---PPQSPYRVSYQQQQQSHHYPGYPPQPOTQYQPGQAYPGPPTQGYGPP-PPGPPNA	501
Qy	153 RRADFNNGLPAGAFSYSDGKPGQWDVPAANGAAPPYKGDSPRTQVVGSGQRRGILPSVPG	212
Db	502 AQGGVHHG-PAGATGAG--HGYQPNAGAGQGGPPPGAYPPPP--GSQQ-----	VPPVPG 551
Qy	213 RATPVTN---GVNGTGKNTTIPAKDADGKF-PCPNCKNTYLHAKHLKRHLRHTGDRPY	267
Db	552 QQQPPPGPPPPGQPPPTGCGQQQPPPPGPPQSQYQPPPPQNS-----AGGPPP	596
Qy	268 MCVLCKDTFSRSDILKRHFQKCSIRRNPTG-----ATHLSHPNAHVKRSQQQAAA	318
Db	597 M-----GYAGYPPNPGQYGAGAGGPPPPSGYWPPTTSSAQSPYQAYQQQQQAAA	650

```
QY 319 N-----PVKPVQDEVST-----VPPNGIPGTTYGEGAVN----- 349
Db 651 GGGAGAPGSSYPGGPPTSGAAPPpPGGAYSTTAPSOtPPpQGGGAGGNNPNPNA 710
QY 350 -----GNGLAPARPGYADHQTmGfPMSSVNGMGRGQPEDAFpGGRPHQG-APWP 397
Db 711 QQSTPPpQGGAGGAGSGPGGAGQYAGPPQ-----QPQQQPPGVVSVGVAPLP 762
QY 398 ---QAPKQSP--YLVQPGADPSGHQLNIDRNIEQVKQPVV----- 432
Db 763 TQVQPTYSTPGNYNQPGAPPPNQOQQOQQOQQOQQOQTTPPSAGSGGGAPNAQOGQ 822
QY 433 -QDP-----KRPVMPGHPGHPGELDTSMFQPOAPEG-----YMFQSMPGGQEPIHA 479
Db 823 NQOPPENGATPPMPFNQYQPA-----PGAPQGYGGPPPPQAYGPPPPGSAYPGA 873
QY 480 HVETER-----KYYPtTTAGQ 495
Db 874 YHQPPQAGGYAQYpPTQGYQ 893

RESULT 13
Q9RX57
ID Q9RX57 PRELIMINARY; PRT; 839 AA.
AC Q9RX57;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein DR0458.
GN DR0458.
OS Deinococcus radiodurans.
OC Bacteria; Deinococcus-Thermus; Deinococci; Deinococcales;
OC Deinococcaceae; Deinococcus.
OX NCBI_TaxID=1299;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=R1 / ATCC 13939 / DSM 20539 / NCIB 9279;
RX MEDLINE=20036896; PubMed=10567266;
RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
RA Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
RA Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
RA Vamathevan J.J., Lam P., McDonald L., Utterback T., Zalewski C.,
RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
RA Fraser C.M.;
RT "Genome sequence of the radioresistant bacterium Deinococcus
RT radiodurans R1.";
RL Science 286:1571-1577(1999).
DR EMBL; AE001904; AAF10038.1; -.
DR PIR; F75518; F75518.
DR TIGR; DR0458; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 839 AA; 79759 MW; 3B6C2CD0CFEFFF4DE CRC64;

Query Match 7.8%; Score 224; DB 16; Length 839;
Best Local Similarity 23.2%; Pred. No. 1.2e-07;
Matches 127; Conservative 45; Mismatches 222; Indels 154; Gaps 23;

QY 7 PSRPPSTSLPQGSAPLPSAPISSMPMPQYT---MQPQYFVSQHTLPLPQHHSQSPAP 62
Db 146 PLKPPVQDTPPPPVTPKPVTPPEVTPKPAPTPPEVLQPPVAQTPPVAKPPV-----PAP 198
QY 63 HSYMGQPPYRPDLNRYPASSHDVYASSAAPIMPTTtVGSLLPPTSLSHPNPQAAQAQQS 122
Db 199 TSQTPTPPVQAPTRTPPPQ---AARPTNAPAQT--PAPATQAPAAQTPTAQAPATQT 252
QY 123 PHYPpHSLVLPASSAQSYPQPIAPAPPRDRADFNGLPSGAFYSdGKPGQGDpVAAN 182
Db 253 PATPAPAAQRP---AGGAPSP-APAPAQ-----ANAPAGSVVBEATVPESSTPAAPS 300
QY 183 G-AAPYPGKDSPrTQTVGS-----QGRRGILPSVpGRATpVTNGVNG- 223
Db 301 AQTPTPTTRETAAQTEASPAAPNSSAAAPNEPASEPVAGRPGTAASSPESASpVTVTPRGE 360
```

```
QY 224 ---TGKNTTIPAKDADGKFPCPCNCKNTYLHAKHLKRHLRLHTGDRPYMCVLCKDTFSRSD 280
Db 361 TPDTAASAGTPSAGRVTPAPAPASASE----- 386
QY 281 ILKRHFQKCSIRRGNTGTATHLSHPNAHVKRSoQQAANPVKPVQDEVSSSTVP-PPNGIP 339
Db 387 -----GASAARTPGAG-----SQTPPIPATPIP 409
QY 340 GTTYGEGAVNGNGLAPARFGYADHQTmGfPMSSVNGMGRGQ--PEDAFP-GGRPHQGAPW 396
Db 410 ATPAGRSSGESAGTAAARFNAAP-APVSEDRSDVSGLPRREDAPAESSPVAASPARGA-- 466
QY 397 PQAPKQSPYLVQPGADP-SGHQLNIDRNIEQVKQPVVQDPKRPVMPGHPGHPGELDTWTSM 455
Db 467 SSAPSSAPAAAAPVSRAPVSGSVsAPRTAPTA--PVAEQGEVPVSPSAAAPRGGA--SSA 522
QY 456 FQQAPE-----GYMFSQSMpGGQEPIHAHVETERKYYPtTTAGQESGMNGLYLA 505
Db 523 AAPSApAAAARGSGAAGGAAGGASAPAAARPAQT-----PGASAGGASG-GGEGVS 572
QY 506 STMSGDGT 513
Db 573 ARPSQGGT 580

RESULT 14
P70200
ID P70200 PRELIMINARY; PRT; 1175 AA.
AC P70200;
DT 01-FEB-1997 (TrEMBLrel. 02, Created)
DT 01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE DRPLA protein.
GN DRPLA.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97224514; PubMed=9070948;
RA Oyake M., Onodera O., Shiroishi T., Takano H., Takahashi Y.,
RA Komonami R., Moriwaki K., Ikeuchi T., Igarashi S., Tanaka H.,
RA Tsuji S.;
RT "Molecular cloning of murine homologue dentatorubral-pallidolusian
RT atrophy (DRPLA) cDNA: Strong conservation of a polymorphic CAG repeat
RT in the murine gene.";
RL Genomics 40:205-207(1997).
DR EMBL; D87744; BAA13450.1; -.
DR MGD; MGI:104725; Drpla.
DR GO; GO:0005737; C:cytoplasm; IDA.
DR GO; GO:0005634; C:nucleus; IDA.
DR GO; GO:0015070; F:toxin activity; IDA.
DR GO; GO:0009404; P:toxin metabolism; IDA.
DR InterPro; IPR002951; Atrophin.
DR Pfam; PF03154; Atrophin-1; 2.
DR PRINTS; PR01222; ATROPHIN.
SQ SEQUENCE 1175 AA; 123642 MW; DEBDCE0FC05CBF6 CRC64;

Query Match 7.8%; Score 222.5; DB 11; Length 1175;
Best Local Similarity 23.4%; Pred. No. 2.3e-07;
Matches 130; Conservative 49; Mismatches 199; Indels 177; Gaps 32;

QY 4 RNHPSRPPSTSLP-----QGSAPL--PSAPISSMPMPQYTMQPYPVSQPHTLPLPQPH 55
Db 253 KQHP--PPTTPIPISSSGASGAPPKAPKPSAPVGGSLP--SAPP--PASFPHTPNLPP- 305
QY 56 HSQSPAPHSYMGQPPYRPDLNRYPASSHDVYASSAAPIMPTTtVGSLLPPTSLSHPNPQA 115
Db 306 -----PPALRFLNNASASPpGM---GAQPIPGHL-----PSPHA 336
QY 116 QAQA-----QQSPHYFPpHSLVLPASSAQSYPQPIAPAPPRDRADFNGLPSGAFS 167
```

```
Db 337 MGQGMGLPGGPEKGTLPASPHPLPPASS-----APGPP--MRYPYSSSSSAAAS 387
QY 168 YSDGKPGQWDPPVAANGAAPYPGKDSPTQV-VGSQGRRGILPSVPGRAT-----PVTN 219
Db 388 SSSSSSSASQYPASQALPSYPHSFPPTSMVSVPKYPQTQSLPSQAVWSQGPDPDPY 447
QY 220 GVNGTGKNTTIPAKDADGKFPCPCNCKNTYLHAKHLKRHLRHTGDRPYMCVLCCKDTFSRS 279
Db 448 G-RLLGNNNTHP-----GFPF-PTGGQSTAHPAATHH---HHQQQPQ----- 485
QY 280 DILKRHFQKCSIRGN-----PTGA-----THLSHPNAHVKRSQQQAAANPVKPVQD 326
Db 486 ---QQH-----HHGNSGAPPPGAYPHPLESSNSHHAHP---YNMSPSLGSLRPYPP--- 530
QY 327 EVSSIVPPNGIPGTTYGEGAVNGNGLAPARPGYADHQTMGFFMSSVNGMGRGQPEDAFP 386
Db 531 -GAHLPPPHG--QVSYNQAGPNP-----PVSSNSSG-SSSQASYS 569
QY 387 GGRPH-----QGAPWPQAPKQSPYLVPQADPSGHQLNIDRNIQVKVPVQDKRPVMP 441
Db 570 CSHPSSSQPGQASYPFP-----VPPVTSSA-----TLSTVIATVASSPAGYKTA 616
QY 442 GHPGHPGELDWTSMFQQAPEGYMFQSMPGGQEPHIAHVETERYKYPTTTAGQESGMNG 501
Db 617 SPGPP-----QYSKRAPSPGSYKTATPPGYKPGSP-----PSFRTGTTPPGYRG 660
QY 502 LVLASTMSGDGTVP 516
Db 661 ---TSPPAGPTFKP 672
```

RESULT 15

```
Q7Z718
ID Q7Z718 PRELIMINARY; PRT; 413 AA.
AC Q7Z718;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Specificity protein 7 short isoform.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Milona M.-A., Gough J.E., Edgar A.J.;
RT "Expression of human specificity protein 7 in osteoblasts.";
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY150674; AAN85557.1; -.
SQ SEQUENCE 413 AA; 43092 MW; FD6F791FFDCB7073 CRC64;
```

Query Match 7.7%; Score 221.5; DB 4; Length 413;
Best Local Similarity 24.1%; Pred. No. 7.6e-08;
Matches 111; Conservative 30; Mismatches 139; Indels 181; Gaps 22;

```
QY 11 PSTSL-----PQGSAPLPSA-----PIS-SMPMPQYTMQPYQVPSQPHTLPLQPHHS 57
Db 49 PFTSTNGLLSPAGSPPAFTSGYANDYPPFHSFPFGTGTQDPLLVPKGHS-----S 100
QY 58 QSPAPHSYMGQPPYRPDLNRYPASSHDVYASSAAPIMPHTTVGSLLPPTSFLSHP----- 111
Db 101 SDCLPSVYTSLDMTHPYGSWYKAGIH-----AGISFGP--GNTPTFWMDMHPGGNWL 151
QY 112 NPQAAQAQ-----SPHYPPPHSVLPPASSAQSYQPIAPAPRRDRADFNNGLP SGAF 166
Db 152 GGQGGDGLQGLTPTGPAQPLNQLP-----TYPSEFAPLNP----- 189
QY 167 SYS DGKPGQWDPPVAANGAAPYPG-----KDSPTQVVGSGQR---RGILPSV 210
Db 190 -----APYPAPHLLQPGPQHVLFPQDVYKPKAVGNSGQLESGGAKPP 231
QY 211 PGRATPVTVNGTGNKNTTIPAKDADGKFC-----P 242
```

```
Db 232 RGASTGGSGGYGGGA-----GRSSCDCPCNQCELERLGAAAAAGLRKKPIHSCHIP 281
QY 243 NCNKTYLHAKHLKRHLRHTGDRPYMC--VLCKDTFSRSDILKRHFQ-----KCSIR 292
Db 282 GCGKVIYGKASHLKAHLRWHITGERPPFVCNWLFCGKRFRSDELERHVRTHTREKKFTCLLC 341
QY 293 RGNPTGATHLS-HPNAHVKRSQQQAAANPVKPVQDEVSSIVPPPNGIPGTTYGEGAVNGN 351
Db 342 SKRFRSDHLSKHQRTHE-----PGPGPPPSG--PKELGEGRSTGE 381
QY 352 GLAPARPGYADHQTMGFFMSSVNGMGRGQPEDAFP GGRPHQ 392
Db 382 EEASQTPR-----PSAS-----PATPEKA-PGGSPEQ 407
```

Search completed: April 1, 2004, 17:48:20
Job time : 49 secs